WARWICK
THE UNIVERSITY OF WARWICK

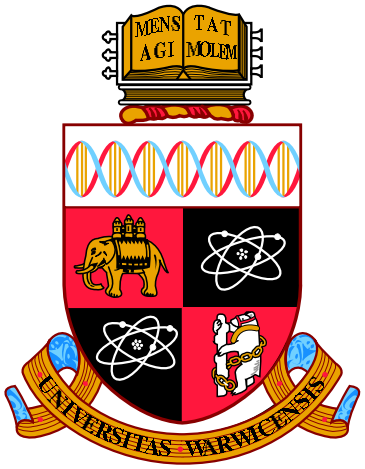
A Thesis Submitted for the Degree of PhD at the University of Warwick

Permanent WRAP URL:
http://wrap.warwick.ac.uk/106777

Copyright and reuse:
This thesis is made available online and is protected by original copyright.
Please scroll down to view the document itself.
Please refer to the repository record for this item for information to help you to cite it.
Our policy information is available from the repository home page.

For more information, please contact the WRAP Team at: wrap@warwick.ac.uk

warwick.ac.uk/lib-publications



Efficient and Adaptable High Dynamic Range Compression

by

Jonathan Hatchett

Thesis

Submitted to the University of Warwick
for the degree of

Doctor of Philosophy in Engineering

WMG

September 2017

Contents

List of Figures

List of Tables

Publications

Jonathan Hatchett, Kurt Debattista, Ratnajit Mukherjee, Thomas Bashford-Rogers and Alan Chalmers. An evaluation of power transfer functions for hdr video compression. *The Visual Computer*, 2016. ISSN: 1432-2315. DOI: 10.1007/s00371-016-1322-0.

Acknowledgements

This thesis would not have been possible without the encouragement, support and kind words given by a great many people, more than I could possibly write down here, but I will do my best.

Firstly, I must thank my supervisors, Alan Chalmers and Kurt Debattista. Alan, for providing unending enthusiasm for me to overcome whatever obstacles lay in the path of my research. Kurt, for providing the honest criticism required to take a collection of unformed ideas and craft them into this work.

I would like to thank Tom Bashford-Rogers and Carlo Harvey for their insightful comments throughout the course of this Ph.D.

My colleagues and dear friends, Josh, Amar and Tim. I'm sorry I made you wait for lunch so many times.

The members of the Visualisation Group: Ratnajit, Demetris, Debmalya, Pınar, Rossella, Martin, Mark, Stratos and Ali, for posing awkward questions during my presentations.

To Savannah, for always believing I would achieve my goal.

To my parents, for providing support and encouragement, a warm bed and a decent meal when I was just about ready to pack it all in. My sister, for unbridled enthusiasm, my brother-in-law, the best academic role-model I could imagine, and to my grandparents for their love.

To the Engineering and Physical Sciences Research Council (EPSRC) for providing the Doctoral Training Grant (DTG) that enabled me to embark on this research.

And finally, to StackOverflow, for proving answers to the perennial question, "why isn't my code compiling?"

Declaration

The work in this thesis is original and no portion of this work has been submitted in support of an application for another degree or qualification at this university or at another university or institution of learning.

Abstract

High dynamic range (HDR) imaging techniques enable the full range of light in a scene to be captured, transmitted and displayed, eliminating under- and over-exposed regions. Storing the full range of light typically requires 32 bits of information per colour channel, four times larger than the 8 bits required for low dynamic range (LDR) data. If HDR is to fulfil its potential in a wide variety of applications such as live broadcast and interactive remote gaming, fast, efficient compression is required to provide HDR video without major changes to existing communications infrastructure. A number of methods have so far been proposed for HDR video compression, however they rely on computationally expensive transformations to either split the video into multiple 8-bit streams or convert to a perceptually uniform domain. This thesis will address the question of whether high-quality HDR video compression can be achieved in a computationally efficient manner by introducing a number of novel techniques.

Initially, the power-law functions used by LDR video are extended to HDR data to provide a straightforward and efficient solution to HDR compression. The Power Transfer Function (PTF) is computationally inexpensive and an objective evaluation shows that it provides improved compression quality at a thirtieth of the computational cost of other leading methods while maintaining equivalent perceived quality.

Subsequently, information about the content and ambient environment at the display are used to adaptively transform the compression method. A subjective evaluation involving 40 participants demonstrates that the necessary peak luminance of content is dependent on the ambient illumination of the display, and an objective evaluation confirms that optimal compression is affected by the peak luminance. An adaptive extension to PTF, Adaptive PTF (PTF_a) is proposed using iterative optimisation to calculate the ideal compression curve, gaining 2.88 VDP over non-adaptive PTF.

Finally, the computational performance of PTF_a is improved by four orders of magnitude to enable real-time compression with little decrease in quality through deep learning. Predictive PTF (PTF_p) utilises a model to predict the compression curved based on the content, display and ambient environment.

This thesis demonstrates a fast, efficient method of general HDR video compression which is extended to provide a fast, adaptive solution for content-specific compression.

Acronyms

AVC Advanced Video Coding

CL Constant Luminance

CRT Cathode Ray Tube

DCT Discrete Cosine Transform

EOTF Electro-optical Transfer Function

GOP Group Of Pictures

HD High Definition

HDR High Dynamic Range

HLG Hybrid Log-gamma

HVS Human Visual System

IEEE Institute of Electrical and Electronics Engineers

ITU International Telecommunications Union

JPEG Joint Picture Experts Group

LDR Low Dynamic Range

LUT Look-up table

MPEG Motion Picture Experts Group

NCL Non-constant Luminance

OETF Opto-electronic Transfer Function

PQ Perceptual Quantizer

PU Perceptually Uniform

QP Quality Parameter

RGB Red Green Blue

SMPTE Society of Motion Picture and Television Engineers

TF Transfer Function

Chapter 1

Introduction

To provide an equivalent perceptual stimulus to the real world has been the goal of imaging since its inception. High dynamic range (HDR) imaging techniques enable the full range of light in a scene to be captured, stored, processed, transmitted and displayed. Without HDR imaging, information can be lost when a scene is captured, manifesting as the under- and over-exposed regions of the image shown in Figure 1.1.

The human visual system (HVS) is capable of differentiating a range of contrast of at least four orders of magnitude at any one moment [BAD+11; RHD+10]. It can adapt however, to lighting conditions which span a range of at least nine orders of magnitude; from dark stars on a moonless night ($1 \times 10^{-3}\,\mathrm{cd/m^2}$) to the sky on a bright, sunny day ($1 \times 10^{6}\,\mathrm{cd/m^2}$) [BAD+11]. Low dynamic range (LDR) imaging technology is limited to capturing and storing a range of at most three orders of magnitude, while HDR imaging can, in theory, represent any range of contrast. Figure 1.2 presents the range that HDR technology can reproduce (at the time of writing).

32 bits of information per colour channel per pixel are typically sufficient to store the full range of light in a scene, four times larger than the 8 bits available for storage with LDR data. Uncompressed, a single, 1080p HDR image contains 24 MB of information [BAD+11]. To fit within the bandwidth constraints of existing communications infrastructure, the high volume of information provided by HDR imaging must be reduced. The study of HDR compression methods is an answer to this problem. Figure 1.3 presents examples of the HDR content evaluated in this thesis.

1.1 Motivation

The traditional computer imaging format is to represent each colour channel, red, green or blue, as a single byte of information. This format is easily portable, quick to compute and each pixel only requires 3 bytes or 24 bits of information to represent. Even with

(a) 30 s (b) 1 s (c) 1/30 s (d) Tone-mapped

Figure 1.1: Stanford Memorial Church with exposure times showing (a) over- and (c) under-exposed regions of an LDR photograph along with (d) a tone-mapped HDR image for comparison [DM97].

an alpha channel to represent opacity, each pixel can fit into a single word on a 32 bit computer. LDR data are typically "gamma encoded" to account for the response of a cathode ray tube (CRT) electron gun to voltage [Poy93]. A side-effect of this encoding is the effect of reducing the visibility of noise or artefacts introduced into the signal. This is achieved via the relation[1], $L = V^{2.4}$, which provides an almost perceptually uniform (PU) encoding over the range of luminances a CRT is capable of emitting, 1 to 100 cd/m^2 [BAD+11]. Even with a PU encoding, the dynamic range LDR imaging is capable of producing is 5.27 stops, assuming a γ of 2.4 and a 5% weber threshold on the visibility of banding artefacts [BC15]. Given that the HVS can simultaneously perceive a range of luminance of around 14 stops [BC15], LDR imaging is only capable of producing a subset of what is perceptible. To capture, store and display a range of information that meets, or even exceeds, that which the HVS can perceive, requires HDR imaging.

Integer precision is not sufficient to store the range of information the HVS is able to perceive, instead decimal data is required. However, fixed precision data is unnecessary as, due to the contrast sensitivity of the HVS the difference between, for example, 10,000.1 to 10,000.2 cd/m^2 is not perceptible while the difference between 0.1 to 0.2 cd/m^2 is. Instead, variable precision, or floating-point, data is more efficient. The format of floating-point data that is most easily computable, IEEE 754 [IEE85], requires 32 bits of information per channel per pixel, therefore a Full-HD RGB image requires 24 MB. If a single image requires 24 MB, then a minute of HDR video at 30 fps requires 42 GB [BAD+11]. With the introduction of UltraHD, comprising 4k resolution and high

[1] Where L is the light emitted, and V is the voltage supplied.

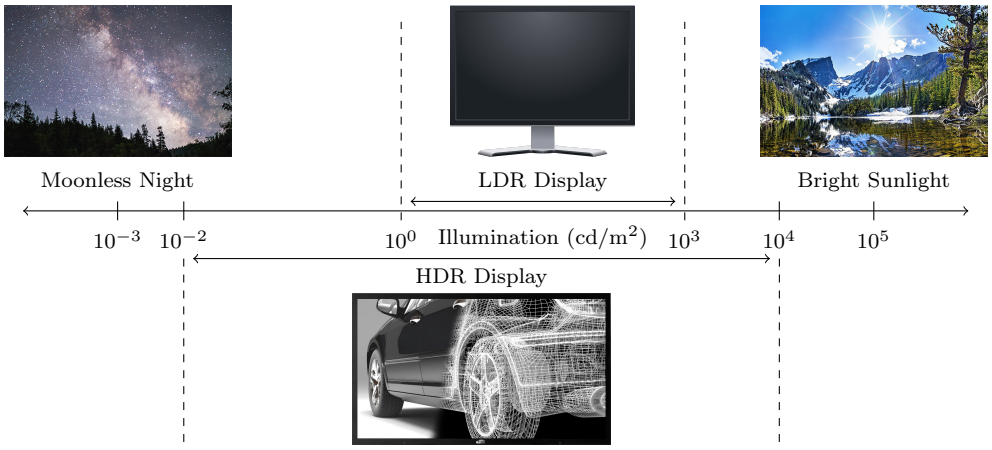


Figure 1.2: Real world luminance levels with commodity displays for comparison [RHD+10, Table 1.1].

(a) Jaguar (b) Seine (c) Fireplace 2 (d) Showgirl 1

Figure 1.3: Tone-mapped HDR content from (a) Warwick, (b) Technicolor [LLF13], (c-d) Stuttgart [FGE+14].

frame rate video, the requirement increases to:

$$\overbrace{3840}^{\text{width}} \times \overbrace{2160}^{\text{height}} \times \underbrace{3}_{\text{RGB channels}} \times \overbrace{4 \text{ bytes}}^{\text{floating-point size}} \times \underbrace{60}_{\text{frames per second}} \times \overbrace{60}^{\text{seconds}} = 358.3\,\text{GB/min}, \tag{1.1}$$

therefore necessitating effective compression.

A number of methods have been proposed for HDR compression. To leverage the coding efficiencies available, they typically rely on pre- and post-processing steps to existing encoding schemes, such as JPEG [Wal92] and AVC [WSB+03]. Traditional encoding schemes were limited to 8-bit data, therefore compression methods [WS06; MEM+06; LK08] were required to split the data into multiple 8-bit streams.

The modern video encoding standards [SOH+12; MBG+13] have enabled the efficient encoding of higher bit-depth information capable of storing HDR video in a single stream [MND13; BC15; EMU16]. This has led to HDR compression finding applications in the film industry [MND13; LFH14] and in broadcast television [BC15] but the solutions provided rely on computationally complex compression functions or memory intensive

look-up tables. The proposed solutions do not necessarily cater for live broadcast, especially in computationally restricted environments such as mobile internet streaming, remote gaming or broadcast hardware. For HDR imaging to fulfil its potential for use in these environments, it is crucial that a computationally efficient codec is available to provide high-quality compression. Improved computational efficiency of HDR video compression would provide other benefits, such as improved battery life for mobile devices or reduced hardware requirements in physical codecs.

During the development and standardisation of current proposals, little attention has been paid towards some of the potential environments in which the content will be consumed. The subjective viewing experiments for the MPEG HDR standardisation [LFH15] have not, so far, been conducted in bright ambient environments. As HDR imagery becomes more widespread, it is likely that the content will be viewed in a wide range of environments, some of which will contain a high ambient illumination. It is important that the increase in detail provided by HDR is perceptible in these environments. To ensure the detail is visible, the luminance of the content may have to be increased, however this may result in degradation of the final image quality. To achieve the highest quality, the most appropriate encoding should be used at all luminance levels at which the content may be viewed . The performance of existing solutions should also be tested in such environments.

1.2 Research Question

Given the need for HDR compression and the computational expense of the current proposed solutions, this thesis will answer the following primary research question:

> *Can the quality of computationally efficient High Dynamic Range video compression be improved by adapting to the environment and display?*

1.3 Methodology

As presented in Section 1.1, current proposals for HDR video compression are computationally inefficient. For HDR video compression to be available on a wide range of platforms, the ideal compression method would be both efficient and capable of producing high-quality encoded video. An evaluation of existing methods in terms of both quality and efficiency should include a new proposed method, designed with both aspects in mind. The evaluation should look at the quality of the decoded HDR content, with reference to the source content. Multiple comparison metrics are available which provide an assessment of image quality, taking a reference image and a distorted image

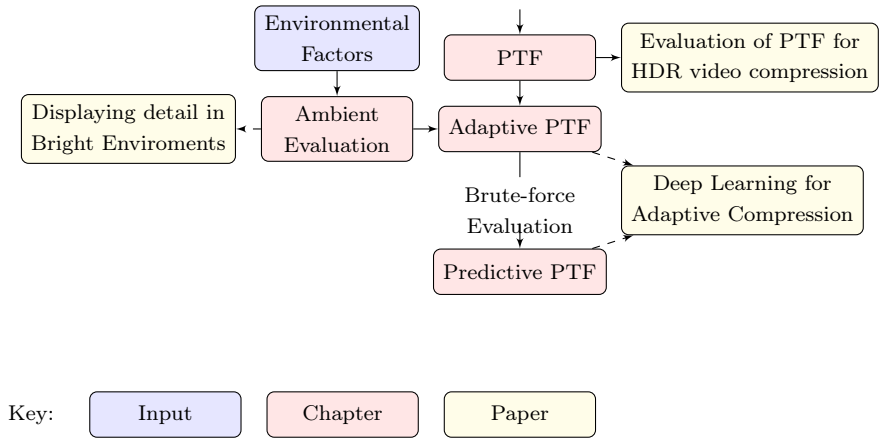


Figure 1.4: The methodology followed by this thesis.

for comparison. It is also important to understand how compression methods perform in a range of scenarios. Different scenarios will likely have different requirements for the quality of the compressed video, therefore the compressed video will be evaluated at a range of quality settings and bit-rates. This will provide a graphical demonstration of how the quality of each method changes with respect to the bit-rate to give an indication of the encoding efficiency. The data on these graphs can be compared numerically to produce pairwise comparisons of the relative efficiency of each method, and also the reduction in bit-rate to achieve the same quality. The computational efficiency of each method should be tested by calculating the duration of time required to compress and decompress a video sequence. As many compression methods are implemented as look-up tables to reduce their computational expense at the cost of an increased memory requirement, an evaluation will compare the performance of the analytic implementations against the look-up tables. The performance of each method can also be presented as the number of video frames per second it can compress and decompress, this will demonstrate whether the proposed method is suitable for use in a real-time environment such as live broadcast. This evaluation is presented in Chapter 3.

If the availability of high-quality, computationally efficient compression is demonstrated, the environments in which HDR content will likely be viewed should be evaluated to ensure that a high dynamic range is perceptible. The subjective evaluation will ask participants to report the dynamic range that can be perceived on a display of a certain luminance at a given ambient illumination. The evaluation will be performed at a range of ambient illuminations that are typical for the locations in which HDR displays will likely be employed. A range of display luminance levels will be tested to correspond with the typical luminance of an LDR display, a current consumer HDR display, a high-end

research HDR display and a prototype, next-generation HDR display. Charts containing an image designed to evaluate the range that is perceptible will be displayed on the screen and each participant asked to report the numbers of chart elements that they are able to see. Each condition of chart, display luminance level and ambient illumination will be viewed by multiple participants to eliminate bias. The results will be reported as the dynamic range that is visible at a given display luminance level and ambient illumination. Analysis of variance (ANOVA) tests will be conducted to ascertain whether either independent variable (luminance level as *brightness*, *illuminance*) has a significant effect on the dependent variable (dynamic range in *stops*). Assuming a significant effect, the results of this evaluation will allow the required luminance to be calculated from the dynamic range of the content and the ambient illumination of the environment. This evaluation is presented in Chapter 4.

If the luminance required to perceive HDR is dependent on the ambient illumination, the ability of HDR video compression methods to adjust to the required luminance should be evaluated. The evaluation will compare the quality of existing methods at a range of luminance levels, tested alongside a new proposed method which adapts to the luminance of the content. The luminance levels chosen will reflect those available across a range of LDR and HDR displays. Content will be adjusted to match the tested luminance levels, and subsequently compressed and decompressed by the evaluated methods. The resulting decoded images will be compared with the adjusted source images using a range of quality metrics to provide an assessment of the compression quality. The metric results will show whether the luminance of the content has an effect on the quality of the compression method. They will also show whether compression performance can be improved by adapting the method to the content and luminance. The computational performance of the proposed method will also be evaluated to decide whether adaptive compression can be used in real-time environment. This evaluation is presented in Chapter 5.

If adaptive compression is required but the performance is not sufficient for use in a real-time environment, then the final step is to optimise the adaption process. The method identified for this is to create a model of the effects of various statistics of the content, such as luminance, and dynamic range, on the compression method required. Multiple techniques exist to create and fit the model, however regression and neural networks have been identified as the most suitable. An evaluation will firstly compare the predictions from the model against the ground-truth data, obtained from the previously proposed adaptive method. If the error between the output of the model and the ground-truth data is sufficiently small then predictions from the model will be evaluated at a range of luminance levels and multiple metrics. The metrics will compare

the luminance adjusted source images with the decoded output from the predicted compression. The metric results will show whether the model is able to accurately predict the required compression to produce high-quality output. The computational performance of the prediction will also be evaluated to ascertain whether predictions can be performed at the rate required for the use of adaptive compression in a real-time environment. If the quality and performance are sufficient then high-quality real-time adaptive compression will be possible in a broadcast environment. This evaluation is presented in Chapter 6. The methodology followed is shown in Figure 1.4.

1.4 Thesis Structure

The structure of this thesis is as follows:

Chapter 2 presents the background to the study of HDR video and the work related to this thesis. The HDR processing pipeline is introduced, and each stage described, including HDR capture; a broad overview of HDR compression techniques including HDR image compression; HDR display including tone-mapping for LDR devices and native HDR display. This chapter discusses the state-of-the-art in HDR video compression and provides descriptions of current techniques.

Chapter 3 introduces the novel contribution of using Power Transfer Functions for HDR video compression. A thorough objective evaluation of PTF is included in which the method performs favourably against the state-of-the-art and a subjective evaluation in which the limitations of the method become clear. This chapter concludes with ideas on how to address the limitations which will be covered in later chapters.

Chapter 4 presents the results of a subjective evaluation to discover the impact of ambient illumination on the perceived detail visible on an HDR display. This evaluation shows that the necessary peak luminance is variable dependent on the ambient illumination and that content should be adjusted to guarantee the entire range is visible. This provides motivation for the chapters that follow.

Chapter 5 proposes a novel method of optimising the compression curve based on the content to provide adaptive compression. The content can be adjusted to provide the luminance necessary to perceive the range contained, or to account for scene-referred content, without an associated loss in compression quality.

Chapter 6 demonstrates that deep learning techniques can be used to train a model to predict optimal values for compression. The resulting model is able to predict

adaptive compression parameters at a rate required for live compression while increasing objective quality over a fixed compression curve.

Chapter 7 provides a summary and conclusion to the thesis and presents suggestions for future work.

Chapter 2

Background & Related Work

This chapter will introduce the concepts required for an understanding of the work presented in this thesis and the current state of the art in this field.

2.1 HDR Imaging

The techniques used in HDR imaging were originally introduced by Ward [War91], Tumblin and Rushmeier [TR93], Mann and Picard [MP94], Debevec and Malik [DM97], and Ward [War94] as a method of recovering the real world radiance of a scene from captured photographs.

2.2 Terminology

A few terms are necessary to understand the content of this thesis however this is not intended to be an exhaustive list of photogrammetric terminology. Banterle et al. [BAD+11] covers photogrammetry in more detail. The terms are defined as follows.

Radiant Power (Φ) is the amount of energy emitted by an object over time, measured in joules per second (J/s) or watts (W). Sometimes known as flux density or luminosity, it is given by the equation:

$$\Phi_e = \frac{\Omega_e}{dt}. \tag{2.1}$$

Radiance (L_e) is the amount of Radiant Power emitted from, travelling through, or received at a point from a given direction, measured in watts per steradian per square meter ($\mathrm{W\,sr^{-1}\,m^{-2}}$), via a particular direction. **Luminance** (L_v) is frequently used instead to denote the amount of Radiant Power received at a point from a given

direction, measured in candela per square metre (cd/m^2), and is often referred to by the term *nit*. Both define the same physical quantity and share the equation:

$$L_{e,v} = \frac{d^2\Phi_e}{dA_e \cos\theta d\omega}. \tag{2.2}$$

Irradiance (E_e) is the amount of energy received on a surface (A), measured in watts per square metre (W/m^2), and is given by the equation:

$$E_e = \frac{d\Phi_e}{dA_e}. \tag{2.3}$$

Range is taken to mean the ratio between brightest and darkest luminance values present. This is also known as contrast ratio (CR).

$$L_{\max} : L_{\min}, \tag{2.4a}$$

$$CR = \frac{L_{\max}}{L_{\min}}, \tag{2.4b}$$

where L_{min} is the minimum luminance value and L_{max} is the maximum luminance value. If L_{min} is 0 then the CR of the image is infinite so typically the first non-zero value is used instead.

Dynamic range (DR) is defined as the logarithm of the range where the log base differs on the field and use.

Stops are the base-2 representation of dynamic range. The unit is most frequently used to measure DR in cinema. It is calculated as follows:

$$\text{stops} = \log_2\left(\frac{L_{\max}}{L_{\min}}\right). \tag{2.5}$$

		Dynamic Range	
Bit-depth	Contrast Ratio	Stops	dB
8	256 : 1	8	48.16
10	1,024 : 1	10	60.21
12	4,096 : 1	12	72.25
14	16,384 : 1	14	84.29
16	65,536 : 1	16	96.33

Table 2.1: The relationship between bit-depth, contrast ratio, stops and decibels.

An increase of 1 stop equates to a doubling of luminance. For example, the difference in luminance from $2\,cd/m^2$ to $16\,cd/m^2$ is 3 stops and the difference in luminance from $32\,cd/m^2$ to $256\,cd/m^2$ is also 3 stops. This exponential growth is useful because it relates closely to how luminance is perceived by the HVS, known as the Weber-Fechner law [Web34; Fec38].

The term stop in imaging is closely related to the term f-stop in photography where it is the ratio of the focal length to the diameter of aperture. On a camera lens it allows control of the amount of light passing through the aperture and on to the sensor. The control of f-stops on a camera is via the iris which closes over the sensor reducing the area by half with each f-stop.

Exposure value (EV) is sometimes used instead of stops as an increase of 1 EV also equates to a doubling of luminance. In photography, the term EV represents a combination of shutter speed and f-stop. Like stops it is dimensionless.

Decibels (dB) can alternatively be used to measure dynamic range as a base-10 representation of the ratio from the maximum value to the minimum value which in imaging systems is peak brightness and the noise level respectively. This term is more commonly used by digital sensor manufacturers. A decibel is defined as follows:

$$\mathrm{dB} = 20 \log_{10} \frac{N_{\max}}{RMS_{\mathrm{noise}}}, \tag{2.6}$$

where $N_{\max}$ is the peak signal value and RMS_{noise} is the root mean squared value of the noise floor. For example if the maximum luminance is $512\,\mathrm{cd/m^2}$ and the root mean squared noise level is $2\,\mathrm{cd/m^2}$ then the image contains $\approx 48\,\mathrm{dB}$ of range or 8 stops. The relationship between bit-depth, stops and decibels is shown in Table 2.1.

2.2.1 Low Dynamic Range, Extended Dynamic Range and High Dynamic Range

Term	Stops
LDR	[0, 8)
EDR	[8, 16)
HDR	[16, ∞)

Table 2.2: Dynamic range terms as defined by MPEG [LFH14].

The imaging systems that are in most frequent use today can typically store 8 bits of data per colour channel, using an additive system to achieve a dynamic range of $2^8 : 1$ or 8 stops. Some new displays can display 10 bit data. The SDI video interface accepts a 10 bit signal and the raw sensor data available from high-end DSLR cameras, such as the Canon 5D Mk.III, works with 14 bit data [Can14].

With the recent push in the field by communities such as the MPEG to increase dynamic range in digital cinema, the terms LDR (also known as SDR) and HDR are gaining the standardised meanings defined in Table 2.2.

2.3 HDR Pipeline

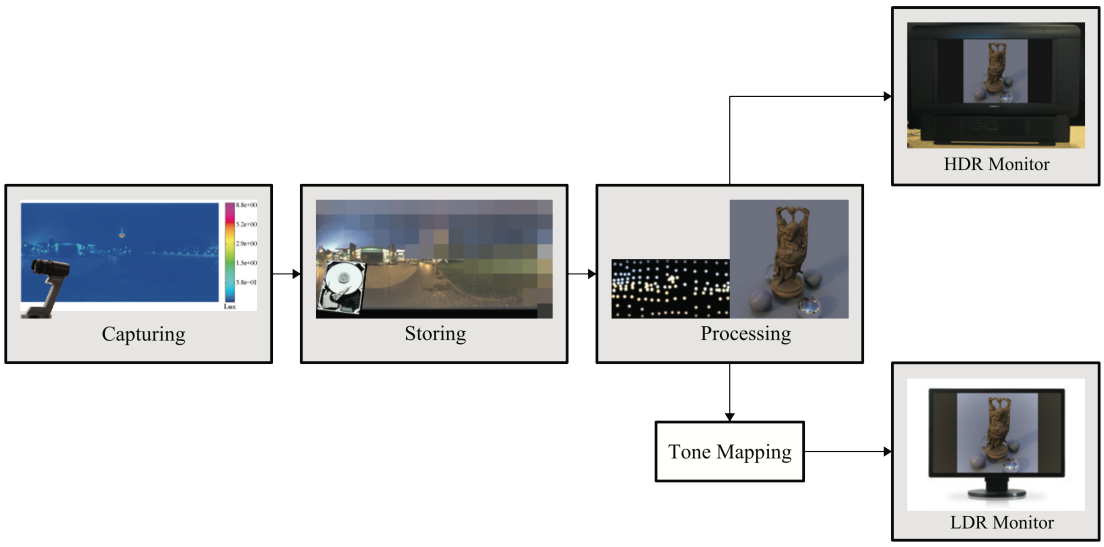


Figure 2.1: The HDR pipeline [BAD+11].

There are four main stages in the HDR pipeline: acquisition, storage and transmission, processing, and display [BAD+11]. Since the problems associated with storage and transmission are related, they can be treated as a single problem. Figure 2.1 shows the pipeline used for HDR imaging.

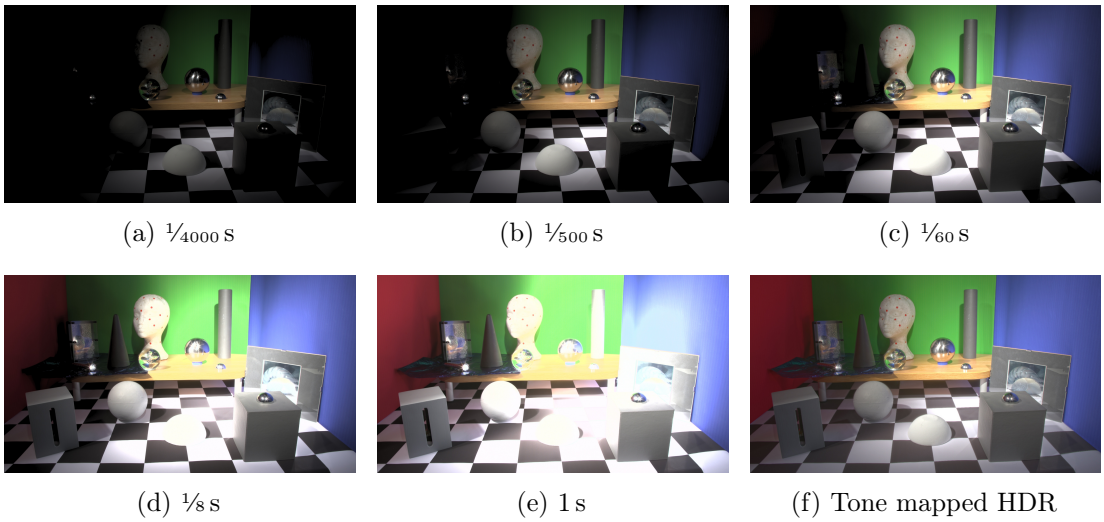
2.3.1 Acquisition

There are multiple ways of acquiring HDR imagery, for example through photography or rendered by computer graphics systems. Each are separate fields of study, however there is some crossover in techniques.

2.3.1.1 Rendering

Computers have been used to create images entirely via algorithms since the 1960's [App68]. The main techniques are rasterisation, ray-tracing and path-tracing, and radiosity. All can generate HDR imagery, however it is path-tracing and radiosity, in their attempt to accurately model the transportation of light, that are placed in a unique position to generate high-quality HDR imagery.

2.3.1.2 Photography



(a) 1/4000 s (b) 1/500 s (c) 1/60 s

(d) 1/8 s (e) 1 s (f) Tone mapped HDR

Figure 2.2: Multiple LDR exposures of the same scene (a-e) for merging into an HDR image (f), tone mapped for presentation [Sel13].

To capture real-world scenes in HDR, novel techniques have been developed as well as specialist and custom hardware.

Mann and Picard [MP94] were the first to publish a method on merging multiple LDR images to create an HDR image. The paper presented the concept of a continuous luminance space as floating-point data in which all of the LDR images can be represented. However the method was rather naïve in its treatment of the camera response function (CRF) providing only a limited method of recovery and with little consideration for image noise [DM97].This is known as multiple exposure capture since the images are usually captured at a range of different brightnesses or EVs. Figure 2.2 shows a sample sequence and the resulting HDR image.

A weighting function, shown in Figure 2.3, is applied to pixels in each image so that only well exposed sections of each image are combined to form the final HDR image. There are drawbacks in combining multiple images to generate an HDR image which are that the scene to be captured must not move in any way in the time when the images are taken otherwise misalignment will occur in the final image. These are known as ghosting artefacts or ghosts.

The transfer function used inside the camera must be obtained in order to gain an accurate ratio between the pixels in each image and produce an output image in which monotonically increasing scene luminance is accurately represented as monotonically increasing image values. This function is known as the CRF since it describes how the camera responds to input luminance. The advantage of doing multiple exposure capture

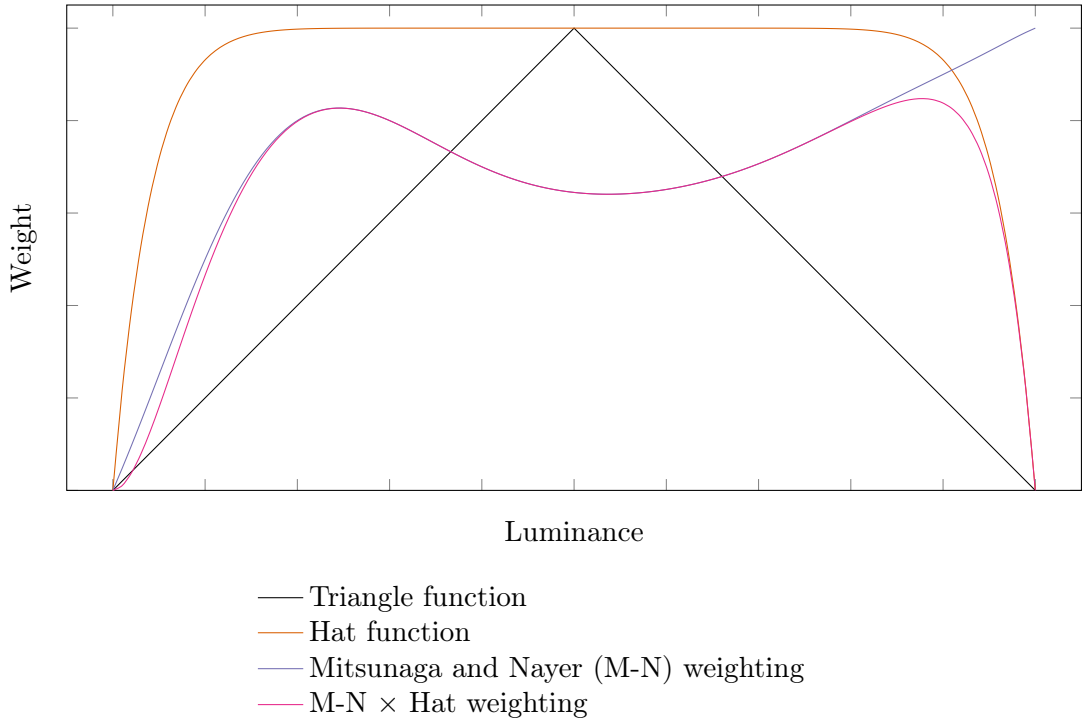


Figure 2.3: Multiple exposure merge weighting functions. Shown are a triangle function, a broad hat function, a typical Mitsunaga and Nayar [MN99] (M-N) weighting function, and the best solution which is a multiplication of the hat function and the M-N weighting function [RHD+10, p. 151].

is that any camera may be used and therefore this method of capture is available to a wide range of people as no special hardware is required, only software to merge the images. This single feature is responsible for the quantity of HDR photography available and has, in a large part, helped the growth of the field in the last few years.

Debevec and Malik [DM97] demonstrated a robust solution for recovering the CRF and calculating accurate radiance from both digital cameras and scanned from film cameras. The method effectively computes the CRF from samples at each pixel and a linear optimisation to minimise the mean squared error and fill in the unknown segments. The weighting function is denoted by $w(z)$, g is the logarithm of the inverse CRF and $\mathcal{O}$ is the objective function in the following equations:

$$w(z) = \begin{cases} z - Z_{min} & \text{for} \quad z \leq \frac{1}{2}(Z_{min} + Z_{max}) \\ Z_{max} - z & \text{for} \quad z > \frac{1}{2}(Z_{min} + Z_{max}) \end{cases}, \tag{2.7a}$$

$$g\left(Z_{ij}\right) = \ln f^{-1}\left(Z_{ij}\right), \tag{2.7b}$$

$$\mathcal{O} = \sum_{i=1}^{N}\sum_{j=1}^{P}\left\{w\left(Z_{ij}\right)\left[g(Z_{ij}) - \ln E_i - \ln \Delta t_j\right]\right\}^2 + \lambda \sum_{z=Z_{min}+1}^{Z_{max}-1}\left[w(z)g''(z)\right]^2, \tag{2.7c}$$

where Z_{ij} is the recorded pixel at image position i and exposure j, E_i is the irradiance value, Δt_j is the exposure time, λ related the smoothness term to the fitting term and g'' is the second derivative of g to ensure smoothness. The paper treats each colour channel as a separate problem however that is rarely the case in colour transformations, however it does yield acceptable results in practice.

Mitsunaga and Nayar [MN99] improved upon the previous work in the field by providing a method of obtaining the true response function with high regard for noise and improved automatic calibration. The method produces a polynomial approximation of the CRF with the advantage of devising the exact exposure ratios. This is useful when the exposure and aperture are not provided by the camera as required by [DM97].

A method of deriving the CRF from the image histograms was presented by [GN02]. The use of histograms instead of patches brings with it the property of matching even with modest scene and camera motion.

Nayar and Mitsunaga [NM00] propose a method of capturing HDR from a single sensor by placing a mask over the sensor in order to reduce transmittance over certain pixels thereby lowering the exposure spatially. This method reduces the resolution of the image in order to increase the DR. It requires a fine filter mesh to be placed in front of the sensor and accurately aligned with the sensing elements. Cho et al. [CKL14] proposed a similar method that utilised the rolling shutter of many cameras to alternate between exposure times, however this too reduces the vertical resolution of the captured data. It is similar to work included in the Magic Lantern firmware for Canon cameras which achieved the same technique by varying the sensor gain for alternate rows of pixels. Another "inventive" method is to search the internet for differently exposed images of the same scene and then align and merge them to boost the DR [JPR14].

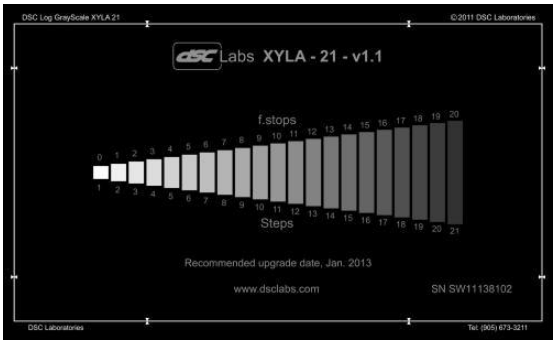


Figure 2.4: DSC Labs Xyla-21 chart [DSC14].

To ascertain the details of the CRF, the captured range can be measured using a

dynamic range chart such as the DSC Xyla 21 [DSC14], shown in Figure 2.4. Panoramic cameras are ideally suited to HDR imagery capture as the wide angle will frequently capture objects in direct sunlight in the same scene as objects in shadow and often the sun itself. Also the long exposure time allowed by panoramic capture means the all objects can be well exposed. The SpheronVR SceneCam [Sph14] is a good example of one such camera and can capture a dynamic range of 26 stops, another is the Civetta 360 which cans capture 28 stops [Wei14].

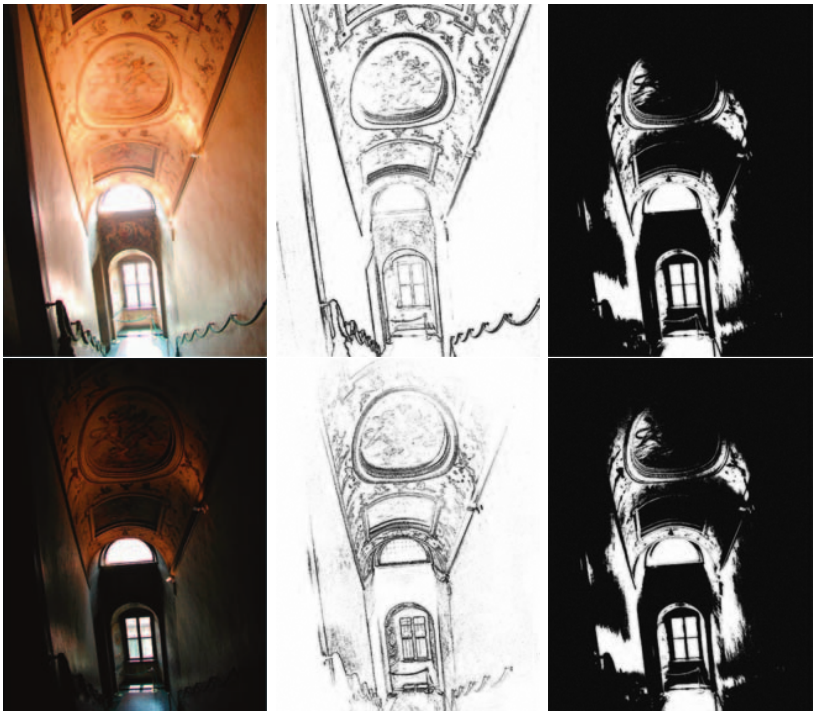
Deghosting is the technique of removing mis-alignment in the image caused by movement between the captured exposures. Two types of ghosting artefacts can occur in multiple exposure HDR imaging, global registration errors and local registration errors. Global registration errors come from a change in the camera position or direction, or camera shake. It can be mostly avoided by using a tripod although sometimes in the creation of video that requirement is too restrictive and other measures must be taken.

Ward [War03] proposed an algorithm to quickly register multiple LDR images before merging to create an HDR image. The idea is to segment each image into what is called a mean threshold bitmap (MTB) by calculating which pixels had a luminance above or below a threshold value, shown in Figure 2.5. The threshold value is set as the median of the overall brightness which gives the algorithm the usual property of exposure invariance. The search for the correct alignment can be done with either brute force within a defined region, gradient descent to find the local minima or by computing an image pyramid which is likely to find the global minima within a defined region. The method is fast because it does not allow for the resampling of the image however it cannot account for rotation between LDR images without modification.

A more robust solution based on MTB was proposed by Grosch [Gro06]. It uses a simplex optimisation of rotation and translation and is similar to a method previously taken by Jacobs et al. [JWL05] but is optimised for use with graphics hardware. In their influential paper Kang et al. [KUW+03] presented a motion estimation alignment technique that could account for both camera and object movement. Based on the work of Lucas and Kanade [LK81], it used an off-line motion vector that was computed between consecutive frames for each pixel.

Local registration errors arise when subsections of the image have moved between each exposure, such as people or cars moving inside the frame. Figure 2.6 shows an example. It can be thought of as two separate problems, ghost detection and ghost removal.

A number of algorithms are available to detect ghosts so they can be removed later. Other techniques treat the problem of ghost detection and removal as one and provide



(a) Individual exposures (b) Edge bitmaps (c) Mean threshold bitmaps

Figure 2.5: (a) Unaligned exposures of an Italian stairwell, (b) incorrectly matched edge bitmaps and (c) correctly matched mean threshold bitmaps [War03].

an integrated solution to the problem.

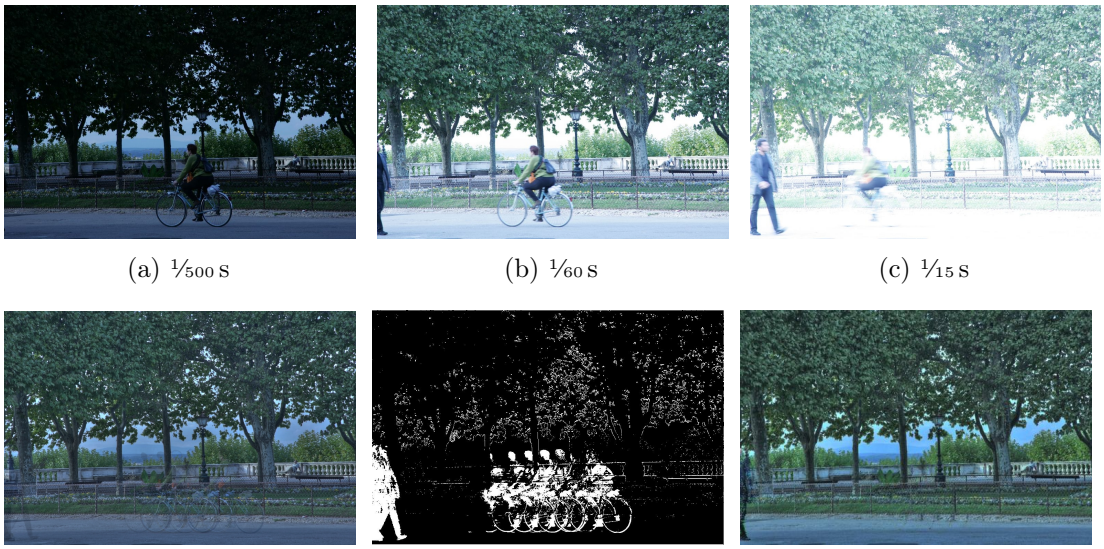
Variance of the exposures can be used to detect ghosts. Regions effected by movement are likely to have a high variance where the variance is taken to be the maximum from the 3 colour channels [SS12]. The ghost map can be obtained by computing a variance image (VI) and then testing it against a threshold [JWL05]. Where the dynamic object and the background are a similar colour the variance based method will fail to detect ghosting. To counter this a local neighbourhood entropy map can be calculated from a local histogram for each LDR image.

Gallo et al. [GGC+09] present a method of detecting ghosted regions using the RANSAC algorithm [FB81]. The method is based on the assumption that the intensity values in 2 images at any location are related by their exposure times. Any location that deviates from the relationship must either be a ghost or is saturated and therefore can be ignored. The algorithm uses patches in order to tolerate noise in the image. The least saturated image is taken to be the reference then RANSAC is used to find the line of best fit through log intensities of the current image plotted against the reference image. The percentage of outliers is calculated using a distance threshold and if the

percentage is higher than a predefined threshold the patch is affected by a ghost. Gallo et al. set the distance threshold to 0.75 and a 40×40 patch is affected by a ghost if the percentage is outliers is greater than 0.5 % of the patch's size.

Joint probability densities can be used to roughly detect ghost regions [HLL+11], then refine the regions using an iterative energy minimisation based on graph-cuts. Like in previous methods joint intensity histograms are created to show the correlation between multiple exposures. After each iteration each join histogram is convolved with a gaussian filter, 5×5 in the method by Heo et al. [HLL+11], and renormalised to represent a probably density function.

Motion estimation can account for both camera movement and local object movement in the frame. Kang et al. [KUW+03] proposed a method which uses motion estimation and gradient-based optical flow to warp pixels from each exposure into a single, consistent image. In their method, global motion between two exposures is first found by estimating an affine transform that remaps an exposure L_k onto exposure L_{k+1}. Then a local motion field is computed via a gradient-based optical flow technique. The difference between L_k and L_{k+1} provides a ghost map however this method implicitly provides both ghost detection and removal. Motion estimation techniques are based on the brightness consistency assumption which does not hold when multiple exposures are used so the LDR images must be transformed into the radiance domain by use of the camera response function. Granados et al. [GKT+13] provide a method of ghost removal by creating a model of the noise in the image and then estimating the likelihood that a pair of colours in differently exposed images are the same irradiance. A single LDR source is used where motion is detected however the method is capable of dealing with a large displacement for objects in the source images where motion estimation would fail. The method works by first aligning the images to remove small camera displacement such as shake. To do this the images are registered using RANSAC from a sparse set of SURF feature-point matches [BET+08]. With the images aligned the method identifies a consistent subset of pixels at each location, much like the variance detection method, based on the noise distribution which is derived from the camera and exposure settings and can be modelled by a gaussian distribution. The use of a consistent subset requires no reference image need be selected and so the algorithm is more accurate at ascertaining the true background for the scene. The method may even generate multiple ghost-free images so a final HDR image is selected so ghost regions do not overlap and the image has the highest signal-to-noise ratio.



(a) 1/500 s (b) 1/60 s (c) 1/15 s

(d) Naïve merging (e) Ghosting detected (f) Ghosts removed

Figure 2.6: (a-c) 3 LDR images from a set of 7 containing an outdoor scene traversed by a cyclist [SS12], (d) naïve HDR merging showing multiple ghosts in the result, (e) the detected ghost map, (f) ghost free HDR image computed by the method described in [SPS+09].

(a) HDRv, courtesy of Spheron VR Ag. (b) CamCube.

Figure 2.7: HDR video cameras.

2.3.1.3 Video capture

The film industry is the main force behind the development of new commercial video cameras. The move from analogue to digital capture caused a reduction in the DR available as early digital camera and storage technology did not allow for a high DR. This left space for prototype camera systems to be developed by small volume manufacturers such as HDRC CamCube [HDR17] and Spheron VR, who's HDRv camera is capable of capturing 20 stops, examples are shown in Figure 2.7.

In recent years the demand for HDR imagery has increased and the commercial camera industry has caught up with research and many camera manufacturers are building cameras that can capture in HDR. The Red Epic [Red14] can capture 18 stops, the ARRI Alexa [Arr17] can capture 14.5 stops, the Sony F65 [Son14] can capture 14 stops, examples shown in Figure 2.8.

(a) Red Epic Mysterium-X (b) Arri Alexa XT (c) Arri Amira (d) Sony F65

Figure 2.8: HDR cinema cameras.

Professional cinema cameras come with a large price tag and therefore, a need for more reasonable cost HDR camera systems is present. The Magic Lantern firmware for Canon EOS cameras [Mag14] can partially fulfil this need. However, there is a trade off to be made for using low cost systems such as Magic Lantern. Although raw data from the sensor is available from the Canon, the 5D Mk. III only achieves ≈ 11.6 stops from 14 bit data showing that the data is expanded to cover all the bits. To improve the DR the camera is capable of capturing, Magic Lantern employs other methods such as multiple exposures and ISO interlacing to boost the DR.

Beam splitting is the process of splitting the light entering the camera into separate beams using semi-transparent mirrors then irradiating multiple standard camera sensors [KGB+14]. Beam splitting cameras do still suffer from alignment issues similar to those found in multiple exposure ghosting however the misalignment is fixed. Algorithms have been developed to fix the alignment issues in real-time [KGB+14]. Another issue with beam split capture is that light is removed from the brightest sensor to provide information for the darker sensors. The reduction in light for the brightest sensor results

in a dark frame overall, potentially hampering low light performance.

2.3.2 Storage

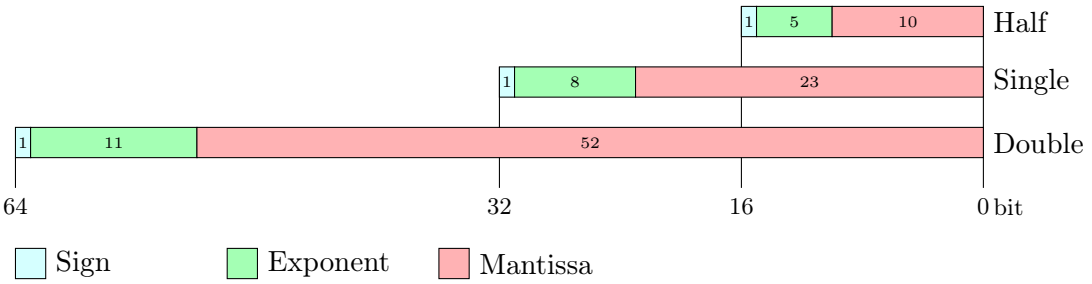


Figure 2.9: Half, single and double precision floating-point bit layouts as described by the IEEE 754-2008 [IEE08] standard.

Once acquired, the digital HDR image must be converted in a format suitable for storage and subsequent distribution. The native representation for processing HDR data is called floating-point. Floating-point is used as the location of the decimal point in the number representation is variable, this gives a fixed number of significant binary digits, but variable precision. IEEE 754 is the standard for floating-point [IEE08] and has 16, 32 and 64 bit variants. The dynamic range that can be represented by the 32 bit variant is $\frac{3.4 \times 10^{38}}{1.18 \times 10^{-38}} \approx 254$ stops. The move from 24 bits per pixel (bpp) for LDR imagery to 96 BPP for full precision HDR imagery results in a 4 fold increase in information stored. This means that a standard RGB 1920×1080 resolution image requires:

$$\overbrace{1920}^{\text{width}} \times \overbrace{1080}^{\text{height}} \times \underbrace{3}_{\text{RGB channels}} \times \overbrace{4\,\text{bytes}}^{\text{floating-point size}} = 24.8\,\text{MB} \tag{2.8}$$

of storage space. The value, n, stored by a floating-point number is given by the formula:

$$n = (-1)^s \left(1 + \sum_{i=1}^{b(m)} m_{b(m)-i} 2^{-i}\right) \times 2^{e-(2^{b(e)-1}-1)}, \tag{2.9}$$

where s is the sign bit, e is the exponent and m is the mantissa and $b(x)$ is the number of bits in x.

The inherent issue with HDR imaging is the storage required. Various compressed HDR image formats have therefore been proposed. For static imagery, the most common of which are OpenEXR and Radiance HDR.

2.3.2.1 Radiance HDR Format

Ward [War91] proposed an image format (.hdr) which became the native format for the Radiance Lighting Simulation and Rendering System [War94]. Each pixel contains 32 BPP comprised of three 8 bit red, green and blue channels plus an 8 bit shared exponent. Alternatively, the format can store three channels containing data in the CIE XYZ colour space plus a shared exponent [War11, p. 28]. Radiance images can encode a DR of $\approx \frac{3.4 \times 10^{38}}{1.18 \times 10^{-38}} \approx 254$ stops using the equations:

$$E = \lceil \log_2(\max(R_w, G_w, B_w)) + 128 \rceil , \tag{2.10a}$$

$$R_m = \left\lfloor \frac{256 R_w}{2^{E-128}} \right\rfloor , \tag{2.10b}$$

$$G_m = \left\lfloor \frac{256 G_w}{2^{E-128}} \right\rfloor , \tag{2.10c}$$

$$B_m = \left\lfloor \frac{256 B_w}{2^{E-128}} \right\rfloor , \tag{2.10d}$$

and can be decoded using the equations:

$$R_w = \frac{R_m + 0.5}{256} 2^{E-128}, \tag{2.11a}$$

$$G_w = \frac{G_m + 0.5}{256} 2^{E-128}, \tag{2.11b}$$

$$B_w = \frac{B_m + 0.5}{256} 2^{E-128}, \tag{2.11c}$$

where $[RGB]_w$ are the original values, $[RGB]_m$ are the encoded mantissas and E is the encoded exponent.

Radiance has the advantage of being ⅔ the size of an OpenEXR image however it loses accuracy if there is not a close correlation between the values in the three channels due to the shared exponent.

2.3.2.2 OpenEXR Format

Industrial Light & Magic, a subsidiary of Lucasfilm, introduced the OpenEXR (.exr) image format in 1999 [Ind14]. Initially aimed at the special effects industry, it has gained significant traction in the field of computer graphics due to its support for HDR images. Initially the half-precision float-point data-type was supported however development on the format has continued and now single-precision float-point and 32 bit integer images are also available. A wide variety of lossless and lossy compression methods are available ranging from basic lossless compression via run-length encoding to wavelet based compression and blocked-based JPEG-like compression for HDR images.

Assuming the half-float depth, each pixel is comprised of red, green and blue channels and will take 48 bits, with 1 sign bit, 5 exponent bits and 10 mantissa bits. This will allow a stored range of $\approx \frac{65{,}504}{6.10 \times 10^{-5}} \approx 30$ stops with a relative error of $< 5\%$ and a step-size of $< 0.1\%$. The precision afforded by EXRs is enough for many applications and they have grown in popularity to become one of the most used HDR image formats. The maximum value that can be stored as a half-precision floating-point is 65,504 limiting OpenEXRs use for absolute graded imagery.

The compression methods available in OpenEXR can be slow to decode making it difficult to use for video sequences unlike other formats such as the PFM.

2.3.2.3 Point Float Map

Point float maps (.pfm) are the most frequently used uncompressed HDR file format. The format is simply an array of IEEE 754 single precision floating-point in a row-major packed format. The data is preceded by a simple header identifying whether the data is monochrome or 3-channel RGB, the resolution of the image, the absolute scale of the data and whether the data is stored in a big-endian or little endian format. An example header is as follows:

```
PF
1920 1080
-1.0
```

where "PF" signifies full colour data, as opposed to "Pf" for monochrome, a resolution of 1920 pixels in the horizontal direction and 1080 in the verticle, and where the absolute value of `-1.0` signifies the data requires no scaling and the negative sign signifies little-endian byte order.

PFMs are presented as they are often required for storing absolute graded images with high precision.

2.3.2.4 LogLUV TIFF Format

The multi-purpose Tagged Image File Format (.tiff) has extensions for the storage of HDR content. The most popular of which, LogLuv, uses the CIE LUV colour space with a logarithm taken of the luminance [Lar98]. 24 and 32 bit variants of LogLuv are available in LogLuv TIFF encoding. However, the 24 bit version is used infrequently as it involves a lookup table to calculate $[uv]_e$. Only the 32 bit version will be explained in detail.

LogLUV separates the input into luminance and chrominance ($Yu'v'$) using the CIE LUV perceptually uniform colour space. 16 bits are allocated for the L_e and 8 bits for

u_e and v_e resulting in 32 BPP. The data is stored using the equations:

$$L_e = \lfloor 256 \left(\log_2 Y + 64 \right) \rfloor , \tag{2.12a}$$
$$u_e = \left\lfloor 410u' \right\rfloor , \tag{2.12b}$$
$$v_e = \left\lfloor 410v' \right\rfloor , \tag{2.12c}$$

and the data is retrieved using the equations:

$$Y = 2^{\frac{L_e+0.5}{256}-64}, \tag{2.13a}$$
$$u' = \left\lfloor \frac{u_e}{410} \right\rfloor , \tag{2.13b}$$
$$v' = \left\lfloor \frac{v_e}{410} \right\rfloor , \tag{2.13c}$$

where $Yu'v'$ are the original CIE LUV values and $[Luv]_e$ are the encoded values. Conversion to $Yu'v'$ is from CIE xyY with the following equations:

$$u' = \frac{4x}{-2x + 12y + 3}, \tag{2.14a}$$
$$v' = \frac{9y}{-2x + 12y + 3}. \tag{2.14b}$$

Conversion to CIE xyY from CIE 1931 XYZ is calculated with the following equations:

$$x = \frac{X}{X + Y + Z}, \tag{2.15a}$$
$$y = \frac{Y}{X + Y + Z}, \tag{2.15b}$$

and conversion from RGB to CIE XYZ is performed with a matrix calculated from the primaries of the RGB colour space.

2.3.2.5 JPEG-HDR

Work has been undertaken to store HDR data in standard JPEG files [WS06]. The data is stored backwards-compatible manner such that when viewed with a standard LDR image viewer the user will see a tone mapped version of the image, whereas when viewed with a JPEG-HDR aware viewer the user will see the same image but with the HDR data visible.

The LDR image data is tone mapped with either a global or a local tone-mapping operator, see Section 2.3.3.1, chosen at the discretion of the user, then the original HDR image is divided by the tone mapped image to create a ratio image as shown in the

equation:

$$RI(x,y) = \frac{L(HDR(x,y))}{L(TM(x,y))}, \tag{2.16}$$

where $RI(x,y)$ is the ratio image at pixel x and y, HDR is the original HDR image, TM is the tone mapped image and L is a function to compute the luminance. The original HDR image can be recomposed by inverting Equation (2.16) to multiply the tone mapped image by the luminance stored in the ratio image.

The ratio image can be further compressed by downsampling and run length encoding. More recently JPEG-HDR was extended to allow for a wider colour gamut [JNT+13] which has be become part of the new JPEG-XT standard [AMR+15].

2.3.3 Display

The last stage in the HDR pipeline is presenting the content to the user. There are two ways of doing this. One is to convert the HDR content into a format that can be displayed on an LDR display, a process known as tone-mapping. The other is to display the HDR content on a display that can natively reproduce information with a high DR.

2.3.3.1 Tone-mapped Display

Tone mapping is an operation that converts the data in an HDR image into an LDR image ready for display. The aim of a good tone-mapping operator (TMO) is to try and represent the characteristics of the original HDR image, such as global and local contrast, high frequency detail, etc., in an LDR image [BAD+11]. Tone-mapping and TMOs were introduced by [TR93] and are defined mathematically as:

$$f(I) : R_i^{whc} \Longrightarrow D_i^{whc}, \tag{2.17}$$

where $f(I)$ is the TMO on image I, w is the width, h is the height and c is the colour channels in the image.

The LDR image is typically comprised of 8 bit integers with the range $[0, 255]$ therefore quantising the data into discrete steps. However, theoretically any depth output can be used with more or less compression taking place. Two classes of TMOs exist: global TMOs apply the same transformation to every pixel using information that is consistent for the whole image, while local TMOs take into account local variations in content to try and retain as much local contrast as possible.

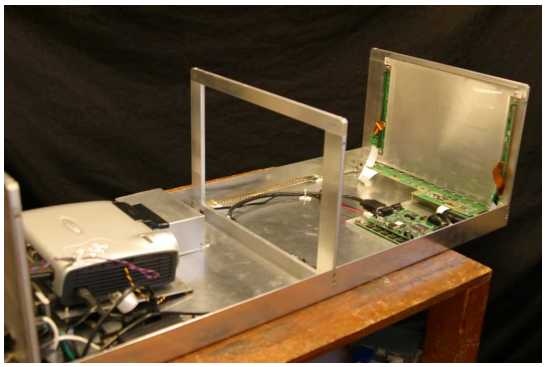
Global Tone-mapping Operators retain the global contrast in the image at the expense of reducing the local contrast. Some TMOs scan the image in a first pass to calculate statistics for how much compression is needed and produce a best-fit function from the input to the output. While other TMOs use the same function irrespective of input and are therefore deterministic with regards to a specific input pixel, hence the output will not be affected by the range in the input and the TMO will be temporally stable. This can be a useful property to have if the tone mapped image is to then be introduced into a temporal structure such as a video. Global TMOs have the property of being invertible so the original range can be retrieved with only the few global statistics used to compress the image needed. The disadvantage of global TMOs is that local contrast and fine detail can be lost due to the quantisation.

Local Tone-mapping Operators retain the local contrast in the image at the expense of reducing or eliminating the broad, low-frequency contrast changes in the image. The image is divided into small blocks and statistics are generated to compress the local content. Contrast between neighbouring pixels is better retained however at object boundaries artefacts such as halos can occur as the local statistics are transferred across edges. Local TMOs are difficult to invert as every pixel would require individual meta-data to restore its luminance. They are also computationally much more expensive as each small region requires statistics to be generated, and introducing temporal stability is difficult.

2.3.3.2 Native HDR Display

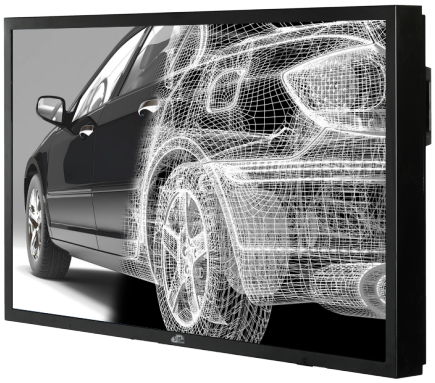
Native HDR requires novel techniques for both processing and display. LDR imaging systems cannot handle HDR without introducing artefacts such as banding and LDR displays do not have the peak luminance or the black level to present a HDR image.

Projection was used to construct the first HDR display. A projector was used to beam light through a front LCD panel. The projector had the colour wheel removed to produce a monochrome beam with a three times increase in luminance. The light was collimated via a lens and then passed through a front LCD panel to provide the final picture. The projector provided the first modulation stage and the LCD panel provided the second. This setup, shown in Figure 2.10a, was able to achieve a minimum luminance of 0.54 $\mathrm{cd/m^2}$ and a maximum luminance of 2,700 $\mathrm{cd/m^2}$ [SHS+04] to give a DR of 12.3 stops.



(a) HDR projector display [SHS+04]

(b) BrightSide DR37-P



(c) Sim2 HDR47E Solar 4K [SIM15]

(d) Dolby Vision reference monitor.

Figure 2.10: Early native HDR displays

An LED Backlit HDR display was first developed by Seetzen et al. [SHS+04]. It was based on an LED back panel (instead of a projector) which passed light through a front LCD panel. 760 LEDs were mounted behind a 18.1" screen however the close proximity of the LED did not cause a problem due to a veiling luminance effect in the HVS [SWW03] which reduces the detail visible in dark areas in close proximity to bright areas because of a scattering effect on light inside the eye. The system produced a maximum luminance of 8,500 cd/m^2 and a minimum luminance of 0.03 cd/m^2 on a checkerboard pattern with squares larger than 20 mm for a DR of 18.1 stops.

The technology developed by Seetzen et al. was developed into a product by BrightSide Technologies, and released as the DR37-P, shown in Figure 2.10b. This display contained 1,380 LEDs behind a 37" LCD panel and produced a maximum luminance of 3,000 cd/m^2 and a minimum luminance of 0.015 cd/m^2 for a DR of 17.6 stops.

BrightSide Technologies was acquired by Dolby in 2007 but the technology was provided to SIM2 and developed by them into the HDR47E Solar 4K [SIM15], shown

in Figure 2.10c, the first commercial HDR display. It supports a resolution of 1080p and using 2,202 LEDs and a 47" LCD front panel is capable of a maximum luminance of 4,000 cd/m^2 and a minimum luminance of 0.4 cd/m^2 for a DR of 13.29 stops. The HDR47E improves upon previous display technology by being air cooled, rather than water cooled in the case of the DR37-P, having a much longer life and by a reduced power usage of < 1.5 kW. It also uses a standard video connection along with a custom data format, based on LogLuv and chroma subsampling, which allows it to be used without any additional custom hardware. SIM2 have increased the maximum luminance capable of being produced by the HDR47E series first to 6,000 cd/m^2 commercially available displays and more recently to a 10,000 cd/m^2 prototype.

With the standardisation of HDR by MPEG, commercial display manufacturers have begun to release consumer displays capable of reproducing HDR imagery. The Samsung JS9800 is capable of emitting 1,500 cd/m^2 at a 4k resolution with a Rec.2020 colour gamut [Sam16]. The Panasonic TX-65DX902B is capable of emitting 1,250 cd/m^2 [Pan17a] while the Sony KD-65ZD9BU is capable of reproducing content at up to 1,180 cd/m^2 [Son17b].

Organic LED (OLED) displays are the final method of displaying HDR content. OLED screens have no backlight to modulate, the light is created and emitted by a front panel of OLEDs. OLED screens are not capable of reaching the same peak brightness as LED backlit HDR displays, they are however capable of reach a lower black level and therefore create a higher DR than LED displays. The Sony KD-65A1 [Son17a] can emit a peak luminance of 1,000 cd/m^2 while the Panasonic TX-65EZ1002B can emit a peak luminance of 665 cd/m^2 [Pan17b].

2.4 Video Compression

There are two broad approaches to HDR video compression: one-stream, and two-stream. Two-stream, or legacy, encoding methods [WS06; MEM+06; LK08] convert HDR data into two, 8-bit streams for encoding with traditional LDR video encoders. These methods typically provide a backwards compatible layer containing tone-mapped information to allow a representation of the HDR content to be viewed in LDR using a standard video player. The second layer contains the residual information needed to combine with an inverse tone-mapped [BDA+09] interpretation of the LDR layer to recontruct original HDR image. A metadata layer is also frequently used to guide the inverse tone-mapping and reduce the size of the residual layer, at the expense of the size of the metadata.

One-stream, or high bit-depth, methods [GT11; MND13; Bor14] utilise the efficient encoding of high bit-depth (> 10-bit) information by more recent video encoding methods [WSB+03; SOH+12; MBG+13]. These methods typically apply a perceptually uniform (PU) encoding to the HDR data either before or after is it converted into a de-correlated colour space before being quantised at the base-band bit-depth and encoded as a video bit-stream. A metadata stream frequently accompanies the video data to enable post-processing software to reverse the PU encoding and scale the HDR video to the appropriate luminance. The metadata is often embedded into the video bit-stream as supplementary enhancement information (SEI) messages, which are provided for in the HEVC standard [SOH+12].

The content provided to the HDR compression method can be one of two formats, scene-referred or display-referred [RHD+10]. Scene-referred information contains the luminance values of the source, whether that is captured, rendered or graded during post-processing. This give the display the maximum flexibility to display the content as true-to-life as possible. The advantage of scene-referred data is that no assumptions about the capabilities of the display are made, and that the information retains many properties of the source environment, providing more information to the viewer about content. The disadvantage is that processing is needed to adjust the data to match what the display is capable of. This could simply be a case of scaling the content to the peak luminance of the display, however this may not fit well with the artistic intend of the content creator. Taking into account the ambient environment at the display could well provide an improved viewing experience at the cost of increased processing.

Display-referred information contains luminance values that should be emitted by the display. The encoded values have no relation to those captured but instead represent the content creators understanding of the luminance that should be seen by the viewer. Little post-processing should occur at the display to account for display peak luminance or ambient conditions. The advantage of display-referred data is that the content creator gets to chose exactly what will be seen by the viewer. However, the disadvantage is that if the display is not one expected by the creator or the conditions in the viewing environment are not ideal, the impression given by the content may be substantially different from that intended.

2.4.1 One-Stream

One-stream, sometimes known as single-stream, HDR video compression methods rely on a transfer function (TF) to create a mapping from real-world luminance values to the integer luma values at the given bit-depth for the video encoder.

Transfer functions typically provide one of two mappings, from real-world values to

encoded values, or from encoded values to real-world values. An opto-electronic transfer function (OETF) takes real-world luminance values as an input and converts them into the values to be passed to the video encoder. An electro-optical transfer function (EOTF) reserves this operation by taking values from the video decoder and converting that to those for display.

2.4.1.1 Perceptual Quantizer

The Perceptual Quantizer (PQ) [MND13] is based on fitting a polynomial function to the peaks in the Barten model for contrast sensitivity [Bar92] as they vary with spatial frequency and over a range of luminances, $L \in 0.001, 0.01, \ldots, 10{,}000\,\mathrm{cd/m^2}$. The method has been put forward as a standard to the Society of Motion Picture and Television Engineers (SMPTE) where it was included in the standard ST.2084 [SMP14]. Since then it has been included in the ITU Recommendation BT.2100 [ITU16] and gained popularity under the term HDR10. It is commonly used in 10-bit and 12-bit variants.

This function can be quantised into a look-up table (LUT) to provide a mapping from luminance to luma for any bit-depth and range of luminance. However, in some applications an analytic implementation is preferred and so a modified Naka-Rushton cone response model [NR66] was found to be a good fit for the LUT.

As PQ is defined as a display-referred method, the forward transfer is the EOTF, decoding the compressed signal for display and is presented in Equation (2.18b). The encoding function is presented in Equation (2.18a).

$$\mathrm{PQ}_{\mathrm{backward}}(V) = \left(\frac{c_2 V^{m_1} + c_1}{c_3 \left(1 + V^{m_1}\right)} \right)^{m_2}, \tag{2.18a}$$

$$\mathrm{PQ}_{\mathrm{forward}}(V) = \left(\frac{V^{\frac{1}{m_2}} - c_1}{c_2 - c_3 V^{\frac{1}{m_2}}} \right)^{\frac{1}{m_1}}, \tag{2.18b}$$

where: $V \in [0, 1]$, $m_1 = \frac{2610}{4096} \times \frac{1}{4}$, $m_2 = \frac{2523}{4096} \times 128$, $c_1 = \frac{3424}{4096}$, $c_2 = \frac{2413}{4096} \times 32$, $c_3 = \frac{2392}{4096} \times 32$.

Both encoding and decoding relations operate in normalised space (0–1) and therefore the incoming signal must be normalised before the TF can be applied. The PQ TF, as standardised in ST.2084, defines a fixed mapping from encoded luma values (known as code word) to display luminance, with the maximum display luminance set to $10{,}000\,\mathrm{cd/m^2}$. This results in unused code words if the peak luminance of content does not reach $10{,}000\,\mathrm{cd/m^2}$. This issue has been addressed by applying a function to re-allocate the number of code words in each segment of the luminance space [LPY+16b].

This has the effect of reshaping the PQ curve. The signal range is divided up in to a number of bins based on their impact to perceptual quality as approximated by the HDR-VDP-2 [MKR+11] quality metric. If a section of the signal contributes little to the final quality, the number of code words in the associated bin is reduced and allocated to bins which contribute more to the quality. To reduce the metadata requirement, the redistribution of code words is translated into a LUT for compression and then into an inverse LUT for decompression. The inverse LUT is approximated by a piecewise polynomial to provide the metadata compression. Since an approximation is used errors may be introduced if there is a mismatch between the desired number of code words per bin, and the actual number calculated by the decoder. To resolve this, the encoder evaluates the piecewise polynomial to generate the inverse LUT, and inverse that to generate the forward LUT expected by the decoder which is then used to perform the re-allocation.

While the implementation of PQ standardised by SMPTE and ITU operated in non-constant luminance (NCL) $Y'C_BC_R$, where the TF is applied to the RGB channels individually, other variations have been proposed. Constant luminance (CL) $Y'C_{BC}C_{RC}$ is supported in ITU Recommendation BT.2020 [ITU15] which improves luma-chroma de-correlation. This crosstalk has the effect of introducing luminance information into the chrominance channel which is then lost when the chrominance is downsampled for compression. Constant luminance colour therefore results in reduced loss in luminance information over non-constant luminance.

An implementation of PQ using a modified intensity, protan & tritan (IPT) colour space [EF98] referred to as IC_TC_P[1] was proposed by Lu et al. [LPY+16a]. The IPT colour space offers improved luma-chroma de-correlation and therefore better constant luminance compared with NCL $Y'C_BC_R$, it also improves hue line linearity and responds better to chroma-subsampling. It was shown to improve results in dE2000 metric [SWD05] by 0.28 over $Y'C_BC_R$ in a coding-only evaluation and reduce the bit-rate required for the same quality by 2.8% when tested with the transfer optical signal to noise (OSNR) XYZ metric [LPY+16a].

2.4.1.2 Hybrid Log-Gamma

Hybrid log-gamma (HLG) [BC15] is a compression method which conforms to the log and gamma segments of the analytic model proposed by Mantiuk et al. [MMS06]. HLG is a scene-referred method, defined relative to a reference luminance taken to be the equivalent of pure white in an LDR capture system, and with a maximum value of twelve times the reference. The gamma segment of HLG approximately matches the tone

[1] The chroma order was reversed, resulting in ITP.

curve of LDR from zero to the reference luminance, affording the method backwards compatibility by presenting this segment alone. Above the reference luminance HLG exhibits a logarithmic curve, with the segments matched at the first differential.

HLG has been proposed as a complete display-independent television system [BC15] with initial standardisation proposed as Arib STD-B67 [RB15]. The method was also included in the ITU Recommendation BT.2100 [ITU16].

The proposed method operates on the scene linear light RGB channels to provide encoded R′G′B′ channels for conversion into NCL Y′$C_B C_R$, Equation (2.19a) presents the encode function and Equation (2.19b) the decode function back to scene luminance. The EOTF to convert from scene luminance to display luminance is presented in Equation (2.20b) and can be applied to each RGB channel using the Equations (2.20d) to (2.20f).

$$\mathrm{HLG}_{\mathrm{encode}}(E) = \begin{cases} r\sqrt{E} & \text{if } E \leq 1 \\ a\log(E-b)+c & \text{if } E > 1 \end{cases}, \tag{2.19a}$$

$$\mathrm{HLG}_{\mathrm{decode}}(E') = \begin{cases} \left(\frac{E'}{r}\right)^2 & \text{if } E' \leq r \\ e^{\left(\frac{E'-c}{a}\right)}+b & \text{if } E' > r \end{cases}, \tag{2.19b}$$

where: $E \in [0, 12]$, $E' \in [0, 1]$, $r = 0.5$, $a = 0.17883277$, $b = 0.28466892$, $c = 0.55991073$.

$$\alpha = L_P - L_B,\ \beta = L_B, \tag{2.20a}$$

$$Y_s = \frac{L_R R + L_G G + L_B B}{12}, \tag{2.20b}$$

$$Y_d = \alpha Y_s^{\gamma} + \beta, \tag{2.20c}$$

$$R_d = R\left(\frac{Y_d - \beta}{12 Y_s}\right) + \beta, \tag{2.20d}$$

$$G_d = G\left(\frac{Y_d - \beta}{12 Y_s}\right) + \beta, \tag{2.20e}$$

$$B_d = B\left(\frac{Y_d - \beta}{12 Y_s}\right) + \beta, \tag{2.20f}$$

where L_P is the display peak luminance, Y_B is the display black level and $L_{[RGB]}$ are the luminance co-efficients for the colour space.

The parameter γ in the above equations corresponds to the system gamma, which is in effect a rendering intent, or OOTF, to provide increased contrast in environments

with a high ambient illumination. It is defined by

$$\gamma = 1 + \frac{1}{5} \log_{10} \left(\frac{L_P}{L_S} \right), \tag{2.21a}$$

where L_S is the luminance of the screen surround.

2.4.1.3 HDRv

One of the earliest was by Mantiuk et al. that extended the existing MPEG-4 encoder and attempted to preserve colour and luminance levels visible to the human eye [MKM+04]. This mapped the real-world luminances from linear RGB to an 11-bit perceptually uniform luma space and chrominance into an 8-bit uniform chromaticity scale similar to that used in LogLUV encoding [Lar98].

2.4.1.4 Adaptive Log-LUV

Garbas and Thoma [GT11] presented a temporally coherent extension of the Adaptive LogLUV function [MT10] suitable for HDR video compression. The proposed method maps real-world luminance into a 12-bit luma space, and preserves chrominance in 8-bit $u'v'$ chroma channels similar to LogLUV [Lar98].

2.4.1.5 Non-linear encoding of HDR

Zhang et al. [ZRB11] developed a method that converts HDR data to a 32-bit LogLUV colour space [Lar98], after which the 16-bit luminance channel is converted to 14-bit by non-linear quantisation, similar to Lloyd-Max optimisation [Sch96].

2.4.2 Two-Stream

Two-stream methods can be considered either backwards compatible or non-backwards compatible based on whether at least one of the streams will present correctly using an LDR video player. Mantiuk et al. presented a backwards-compatible method which, following the overall method proposed by Ward and Simmons [WS04], tone-maps the HDR data to create the backward compatible image [MEM+06]. This image is then restored to a colour space compatible with the original and the difference in luminance between the reconstructed frame and the original, is taken and stored a residual data stream. The decoding is performed by reconstructing the tone mapped image and then applying the residuals previously created. The method proposed by Lee and Kim [LK08] also follows the structure proposed by Ward and Simmons. In this method the backwards compatible frames are generated using a temporally coherent tone mapper and the

residual is created by taking the logarithm of the division of the reconstructed image and the original HDR image. To reduce noise the residual stream is cross-bilaterally filtered [ED04]. Other proposed two-stream methods include goHDR [BAD+10] and optimal exposure [DBS+15].

Mai et al. proposed a backwards-compatible two-stream compression method comprising an LDR layer and a residual layer [MMM+11]. The residual layer is used to reconstruct the HDR frames after the LDR layer has been inverse tone-mapped. The proposed method optimises a piecewise-linear tone-curve in order to provide the most efficient tone-mapped representation of the HDR by minimising the mean squared error (MSE) between reference and distorted HDR frames. The authors then go on to propose a closed-form solution to their method by building and applying a distortion model however they do not consider the associated rate cost of the LDR layer [GRG+17]. Gommelet et al. address this problem in their work by adding a constraint to the perceptual quality of the SDR layer with the intention of preserving the artistic intent of the content [GRG+17]. They propose a new distortion model based on image gradient with the aim of better approximating the rate-distortion performance of video encoders. However, both of these methods are primarily suited to backwards-compatible HDR video compression.

2.5 Metrics

The following three metrics are used to provide results for the evaluation.

2.5.1 PSNR

Peak Signal to Noise Ratio (PSNR) is one of the most widely used metrics for comparing processed image quality. To adapt the method for HDR imaging, L_{peak} was fixed at $10{,}000\,\text{cd/m}^2$ and the result was taken as the mean of the channel results.

$$\text{PSNR}_\lambda = 20\log_{10}\left(\frac{L_{\text{peak}}}{\sqrt{\text{MSE}_\lambda}}\right) \tag{2.22}$$

2.5.2 puPSNR

Perceptually Uniform PSNR (puPSNR) was proposed as an extension to PSNR such that it is capable of handling real-world luminance levels without affecting the results for existing displays [AMS08]. The proposed metric maps the range 1×10^{-5} to $1\times10^{8}\,\text{cd/m}^2$ in real-world luminance to values that approximate perceptually uniform

values derived from a CSF. It is from the remapped luminance that the PSNR is calculated.

2.5.3 HDR-VDP-2.2.1

HDR Visual Difference Predictor (HDR-VDP) is an objective metric based on a detailed model of human vision [MKR+11]. The metric estimates the probability at which an average human observer will detect differences between a pair of images in a psychophysical evaluation. The visual model used by this metric takes several aspects of the human visual system into account such as intra-ocular light scatter, photo-receptor spectral sensitivities and contrast sensitivity. HDR-VDP-2.2.1 has been shown to be the objective metric that correlates most highly with subjective studies [HBP+15; MDB+16a].

2.6 GPGPU

Modern graphics processing units (GPUs) are in essence programmable, parallel, floating-point processors. They are able to execute parallel floating-point instructions with high throughput. Although designed originally to compute raster graphics for display they can be harnessed to compute general purpose workloads instead, this has given rise to General Purpose programming on Graphics Processing Units (GPGPU). Due to the parallel nature of rendering, the GPU has been optimised for tasks that require the same operations to be executed on a large data set and so problems that can be modelled in this way gain the most from being run on a GPU. Representing algorithms as parallel tasks is not necessarily trivial as many program structures and idioms frequently used in CPU programming can no longer be used. Algorithms have to undergo what is called parallelisation. Originally GPGPU had to be undertaken by using regular graphics API such as OpenGL. However in time, specific GPGPU programming languages were written, initially proprietary languages such as Nvidia's Cg and CUDA. More recently standard, cross-platform languages have been released such as OpenCL which supports all major graphics card vendors.

Chapter 3

Extending Power Functions for HDR Video Compression

HDR video compression methods may be classified as either a one-stream or two-stream approach [CMH+15], referring to the number of discrete video bit-streams that are contained. A two-stream method separates the single HDR video input stream into base and detail streams which are then compressed separately according to their individual characteristics. One-stream methods, on the other hand, take advantage of the higher bit-depth available in modern video codecs [WSB+03; SOH+12; GRH16]. A transfer function (TF) is used to map the HDR video input stream to a single, high bit-depth stream and, depending on the method, metadata to aid post-processing before display. A number of the proposed one-stream methods [Bor14; MND13] use complex TFs, requiring many floating-point operations for both compression and decompression. If HDR is to gain wide acceptance, and find use in broadcast, internet streaming, remote gaming, etc., it is crucial that computationally efficient encoding and decoding is possible.

This chapter will evaluate whether straightforward power functions, with their associated computational benefits, can be used to efficiently compress HDR video. It will propose -an HDR video compression method, the Power Transfer Function (PTF), which aims to provide real-time HDR video encoding without a loss in quality or compression performance. It will answer the following research question:

> *Are computationally complex Transfer Functions required for high-quality HDR video compression?*

3.1 Related work

HDR video compression methods can be split into two broad categories: one-stream and two-stream. Two-stream methods have the advantage that they can work well on existing 8-bit infrastructure. One-stream methods, on the other hand, require at least 10-bit infrastructure. The advantage of one-stream methods is that they follow a similar pipeline to those used for LDR video, without the need for secondary streams to be transmitted or combined before display.

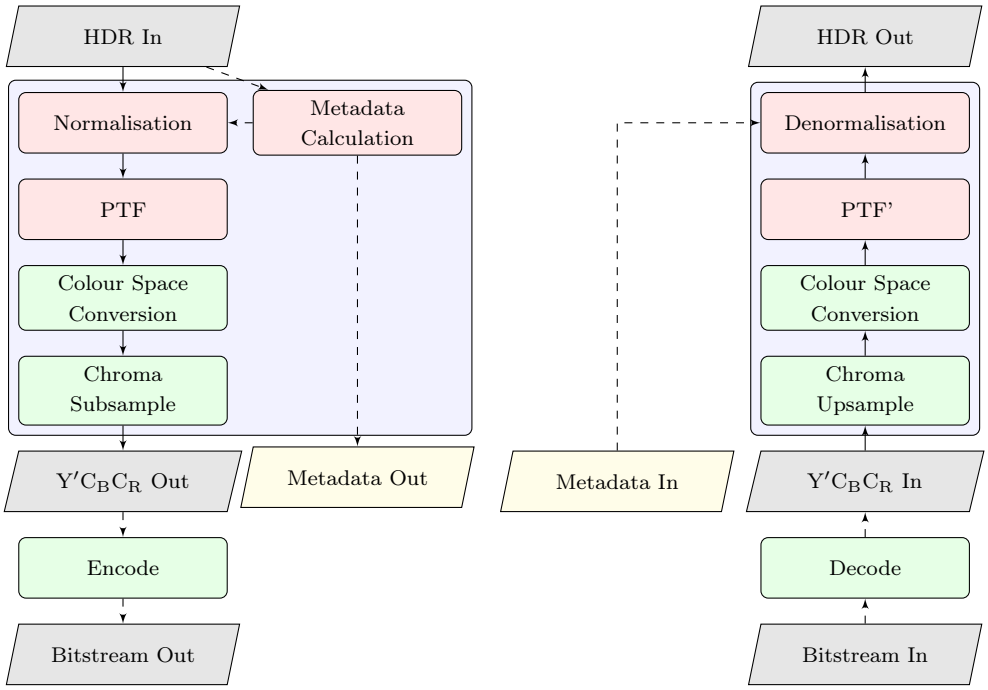
Two-stream methods can be considered either backwards compatible or non-backwards compatible based on whether at least one of the streams will present correctly using an LDR video player. Mantiuk et al. presented a backwards-compatible method which, following the overall method proposed by Ward and Simmons [WS04], tone-maps the HDR data to create the backward compatible image [MEM+06]. This image is then restored to a colour space compatible with the original and the difference in luminance between the reconstructed frame and the original is taken and stored as a residual data stream. The decoding is performed by reconstructing the tone mapped image and then applying the residuals previously created. The method proposed by Lee and Kim [LK08] also follows the structure proposed by Ward and Simmons. In this method the backwards compatible frames are generated using a temporally coherent tone mapper and the residual is created by taking the logarithm of the division of the reconstructed image and the original HDR image. To reduce noise the residual stream is cross-bilaterally filtered [ED04]. Other proposed two-stream methods include goHDR [BAD+10] and optimal exposure [DBS+15].

Several one-stream HDR video compression methods have been proposed in the last 10 years. One of the earliest was by Mantiuk et al. that extended the existing MPEG-4 encoder and attempted to preserve colour and luminance levels visible to the human eye [MKM+04]. This mapped the real-world luminances from linear RGB to an 11-bit perceptually uniform luma space and chrominance into an 8-bit uniform chromaticity scale similar to that used in LogLUV encoding [Lar98]. This method will be referred to as HDRV for the remainder of this chapter. Garbas and Thoma [GT11] presented a temporally coherent extension of the Adaptive LogLUV function [MT10] suitable for HDR video compression. The proposed method maps real-world luminance into a 12-bit luma space, and preserves chrominance in 8-bit $u'v'$ chroma channels similar to LogLUV [Lar98]. This method will be referred to as Fraunhofer for the remainder of this chapter. Zhang et al. [ZRB11] developed a method that converts HDR data to a 32-bit LogLUV colour space [Lar98], after which the 16-bit luminance channel is converted to 14-bit by non-linear quantisation, similar to Lloyd-Max optimisation [Sch96].

The Perceptual Quantizer (PQ) method is based on the fitting of a polynomial function to the peaks in the Barten model of visual perception [Bar92]. Compression is provided by means of a closer fit to a human visual response curve [MND13]. This method has recently been included in a SMPTE standard, ST.2084 [SMP14].

More recently, Borer and Cotton [BC15] proposed a compression method based on the log and gamma segments of Mantiuk's analytic model [MMS06]. This Hybrid-Log-Gamma (HLG) method has been developed to provide support for a display independent television system [BC15], and has also been included in the Arib STD-B67 standard [RB15].

3.2 Power Transfer Function



(a) PTF Encode. (b) PTF Decode.

Figure 3.1: Example pipelines used for encoding and decoding HDR using the PTF. (a) takes in HDR video frames in either scene or display-referred scale and outputs $Y'C_BC_R$ for encoding with a standard encoder. (b) takes as input the encoded bitstream and outputs HDR frames at the initial scale. The dashed lines denote optional processing.

The human visual system (HVS) has greater sensitivity to relative differences in darker areas of a scene than brighter areas [Fec38; Web34]. This non-linear response can be generalised by a straightforward power function. The Power Transfer Function (PTF)

weights the use of the values available to preserve detail in the areas of the HDR content in which the HVS is more sensitive. PTF therefore allocates more values to the dark regions than to the light regions. The theoretical properties of the power functions used in PTF will be presented in Section 3.2.3.

3.2.1 Motivation

The recent addition of higher bit-depth support to commonly used video encoding standards such as Advanced Video Coding (AVC) [WSB+03], High Efficiency Video Coding (HEVC) [SOH+12] and methods such as VP9 has diminished the need for two-stream methods. Instead, this support has motivated an investigation into the efficient mapping of HDR data into 10 and 12 bits. For this purpose, PQ [MND13] uses a perceptual encoding to map the contrast sensitivity of the HVS to the values available in the video stream. This perceptual encoding however, relies on the complex transfer functions shown in Equation (3.1).

$$\mathrm{PQ}_{\mathrm{backward}}(V) = \left(\frac{c_2 V^{m_1} + c_1}{c_3 \left(1 + V^{m_1}\right)} \right)^{m_2}, \tag{3.1a}$$

$$\mathrm{PQ}_{\mathrm{forward}}(V) = \left(\frac{V^{\frac{1}{m_2}} - c_1}{c_2 - c_3 V^{\frac{1}{m_2}}} \right)^{\frac{1}{m_1}}, \tag{3.1b}$$

where: $V \in [0,1]$, $m_1 = \frac{2610}{4096} \times \frac{1}{4}$, $m_2 = \frac{2523}{4096} \times 128$, $c_1 = \frac{3424}{4096}$, $c_2 = \frac{2413}{4096} \times 32$, $c_3 = \frac{2392}{4096} \times 32$.

HLG [BC15] uses a coding-scheme based on luminance relative to peak white for LDR content to compress the HDR data. However this mapping, shown in Equation (3.2), is also not straightforward in nature.

$$\mathrm{HLG}_{\mathrm{encode}}(E) = \begin{cases} r\sqrt{E} & \text{if } E \leq 1 \\ a \log(E - b) + c & \text{if } E > 1 \end{cases}, \tag{3.2a}$$

$$\mathrm{HLG}_{\mathrm{decode}}(E') = \begin{cases} \left(\frac{E'}{r}\right)^2 & \text{if } E' \leq r \\ e^{\left(\frac{E'-c}{a}\right)} + b & \text{if } E' > r \end{cases}, \tag{3.2b}$$

where: $E \in [0,12]$, $E' \in [0,1]$, $r = 0.5$, $a = 0.17883277$, $b = 0.28466892$, $c = 0.55991073$.

This chapter investigates whether a transfer function implemented using straightforward power functions can provide an efficient mapping. Power functions also provide computational benefits, particularly for lower integer powers. To use the PQ transfer function [MND13] requires many calculations, and the HLG transfer function [BC15] has a branch instruction limiting its calculation speed in pipelined systems. However, a

power function can be computed with a single calculation.

3.2.2 Method

Algorithm 1 Power Transfer Function Encoding

procedure $\text{PTF}_\gamma(frames_{in}, \mathfrak{N})$
  for $i \leftarrow 1$,LENGTH$(frames_{in})$ **do**
    $S \leftarrow frames_{in}[i]$
    $L \leftarrow {}^{S}/_{\mathfrak{N}}$
    $V \leftarrow L^{1/\gamma}$
    $Q \leftarrow$QUANTISE(V)
    $frames_{out}[i] \leftarrow Q$
  end for
  return $frames_{out}$
end procedure

Algorithm 2 Power Transfer Function Decoding

procedure $\text{PTF}'_\gamma(frames_{in}, \mathfrak{N})$
  for $i \leftarrow 1$,LENGTH$(frames_{in})$ **do**
    $Q \leftarrow frames_{in}[i]$
    $V \leftarrow$DEQUANTISE(Q)
    $L \leftarrow V^\gamma$
    $S \leftarrow L \cdot \mathfrak{N}$
    $frames_{out}[i] \leftarrow S$
  end for
  return $frames_{out}$
end procedure

PTF is a single stream method, converting HDR input into a single set of compressed output frames. To achieve this compression, PTF utilises the power function: $f(x) = Ax^\gamma$ where: A is a constant, x is normalised image data contained by the set $[0, 1] \subset \mathbb{R}$ and $\gamma \in \mathbb{R}^+$. In Algorithms 1 and 2 the power function and its inverse are expressed by the relations $V \leftarrow L^{1/\gamma}$ and $L \leftarrow V^\gamma$ respectively. PTF can be applied to either the scalar RGB channels or the scalar luminance channel.

The straightforward nature of the PTF method is shown in Figures 3.1a and 3.1b which present the general pipeline into which PTF is used, and from Algorithms 1 and 2 which detail the compression and decompression procedures, PTF_γ and PTF'_γ,

respectively.

Before a HDR video is compressed using PTF, it is normalised to the range $[0, 1]$ with a normalisation factor $\mathfrak{N}$ using the relation $L = {}^{S}/_{\mathfrak{N}}$ where: S is full range HDR data. If the footage is of an unknown range then it can be analysed to determine the correct $\mathfrak{N}$ for encoding, or for live broadcast, $\mathfrak{N}$ can be set to the peak brightness the camera is capable of capturing or the display is capable of presenting.

If the normalisation factor is variable, then it can be stored as metadata along with the video data to correctly rescale the footage for display. Each input frame may be normalised independently, however this may introduce artefacts as the scaling and nonlinearity can interact and lead to the accumulation of errors when using predicted frames. More often a global or temporal normalisation factor is used. The metadata can either be passed at the bitstream level, i.e. with supplemental enhancement information (SEI) messages, or at the container level, i.e. MPEG-4 Part 14 (MP4) data streams.

PTF can be in either a non-constant luminance (NCL) or a constant luminance (CL) encoding system. For use in an NCL system, following compression with PTF, the data must be converted into the output colour space to be passed to the video encoder, and if chroma subsampling is to be used, reduced to the correct format. For use in a CL system, the colour space conversion is first performed on the linear RGB data before the luminance channel only is compressed with PTF. The evaluation in this chapter was performed with NCL encoding.

3.2.3 Theoretical analysis

Figure 3.2 presents a comparison of Just Noticeable Difference (JND) characteristics from various methods and standards. The Greyscale Display Function (GDF) is an implementation of the Barten Contrast Sensitivity Function (CSF) [Bar92] developed for the Digital Imaging and Communications in Medicine (DICOM) standard [AR+98]. The DICOM standard GDF is defined with a lower bound of $0.5 \times 10^{-1}\,\mathrm{cd/m^2}$. The GDF CSF plots a relationship between luminance and luma such that the contrast steps between each consecutive luma value are not perceptible. Methods with contrast steps larger than that of the GDF are likely perceptible at that luminance. Also worth noting is that as the Fraunhofer method is based on log luminance it exhibits a purely linear plot in the logarithmic Contrast Step scale shown on Figure 3.2.

To understand how power functions could be adapted for HDR video compression, the JND characteristics of PTF with the γ values 4 and 8 were investigated. Integer values were chosen as they are expected to exhibit reduced computational cost over non-integer values. Section 3.3.2 will investigate the role of γ in PTF compression and explain why the γ values of 4 and 8 were chosen.

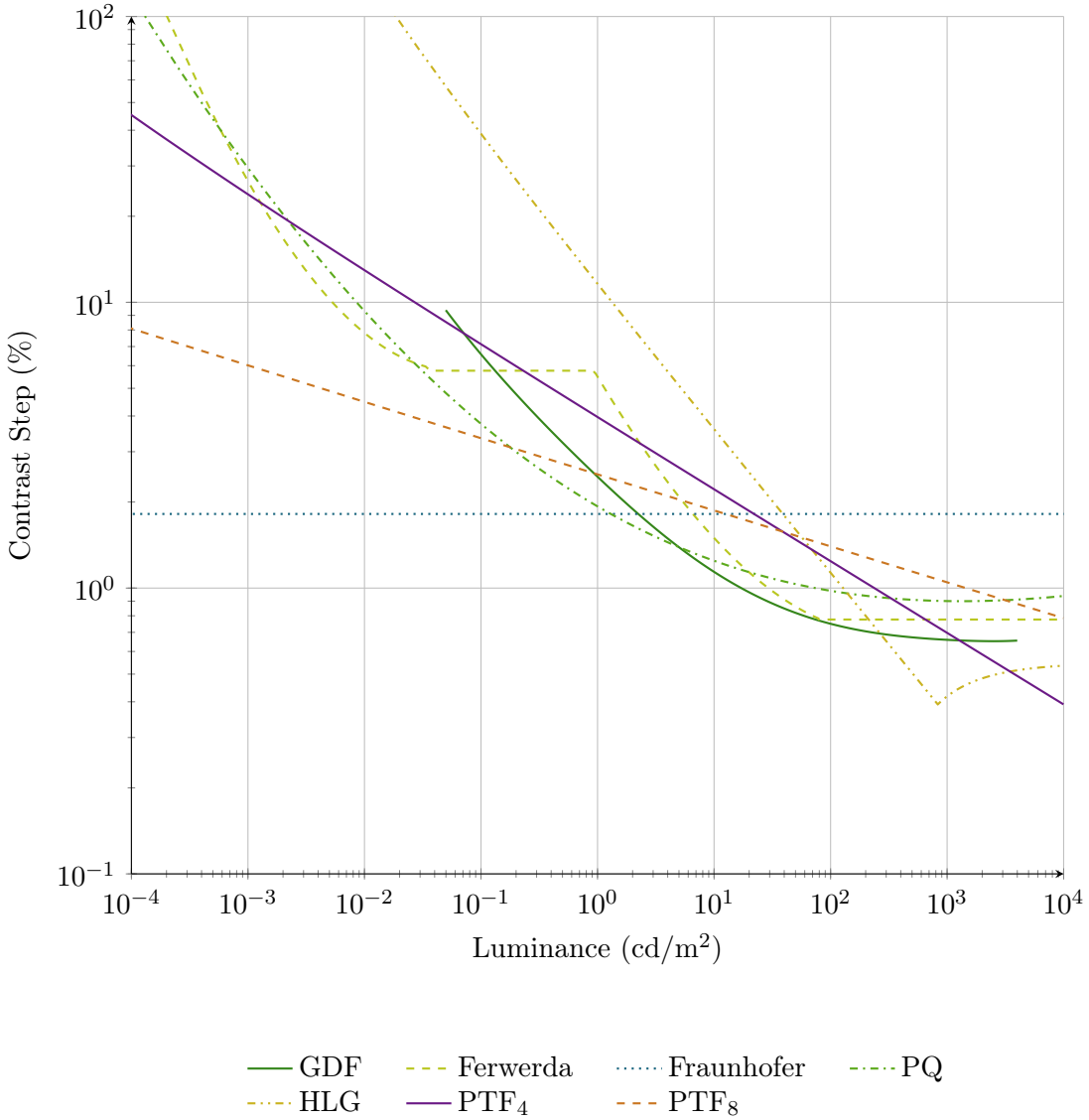


Figure 3.2: Contrast Sensitivity plots showing GDF as implemented by the DICOM standard, Ferwerda TVI [FPS+96] as used by HDRV, Adaptive LogLUV used by the Fraunhofer method, Dolby PQ, BBC HLG and PTF_4 and PTF_8. Luminance and Contrast are shown on logarithmic scales.

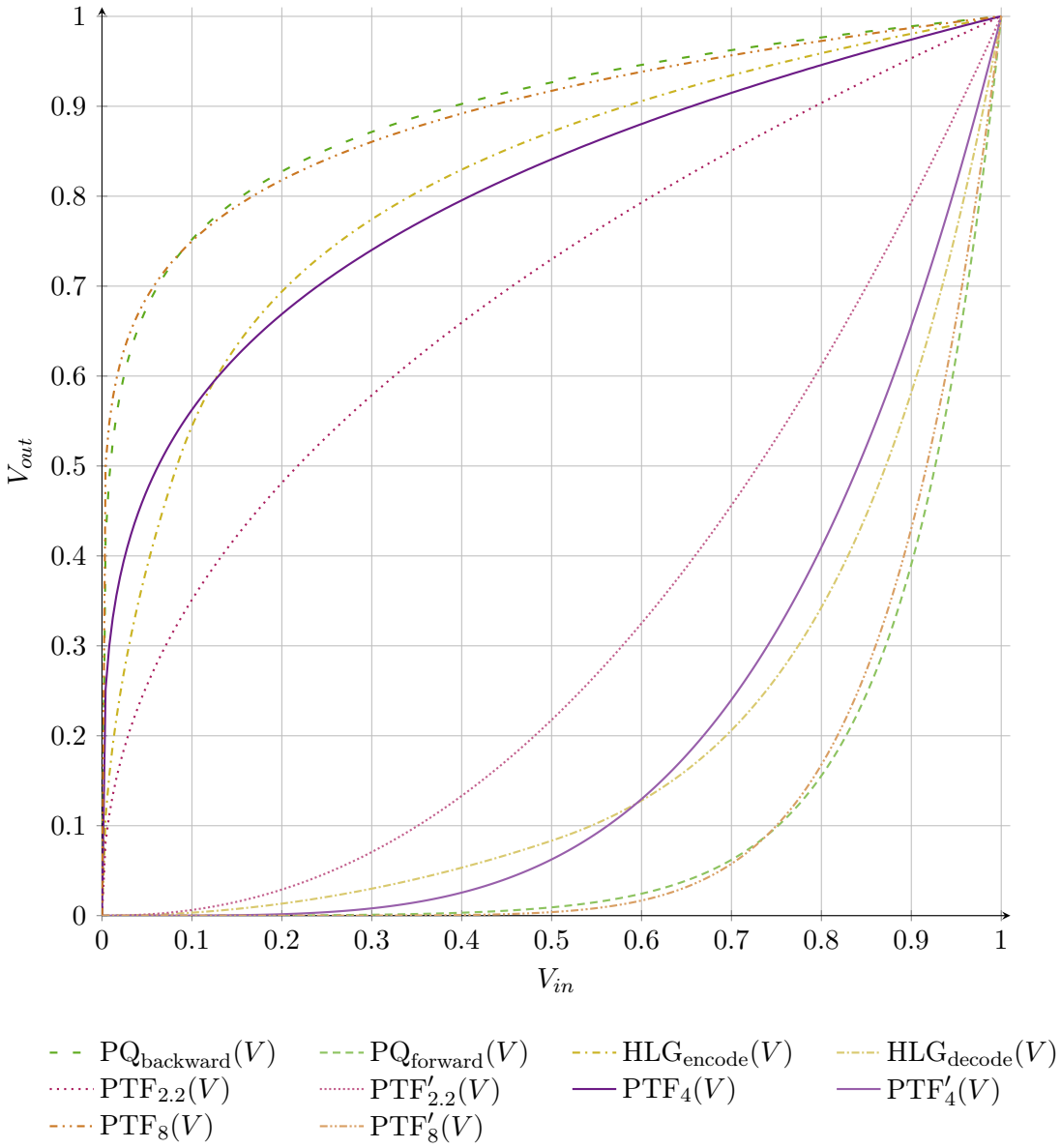


Figure 3.3: Graph showing encoding and decoding transfer functions. Presented are PTF_4 and PTF_8 alongside the PQ and HLG curves. $\mathrm{PTF}_{2.2}$ is presented for comparison with an example LDR Gamma function. HLG has been rescaled to the $[0, 1]$ range for comparison with other TFs.

The power function in PTF is related to the Gamma function used in LDR video systems such ITU-R BT.601 [ITU11] where the effective γ is 2.2, caused primarily by the response of the electron gun [Poy93]. This non-linearity is also perceptually uniform, closely fitting the ½ power law of the de Vries-Rose relationship [Kel77] while providing a Weber fraction [Web34] of $\approx 1\%$ over luminances up to $100\,\mathrm{cd/m^2}$ that a cathode ray tube (CRT) is capable of producing [Poy93]. PTF is more similar to the empirically derived ⅓ power law [SYD87].

Figure 3.3 presents a comparison of the shape of the proposed TFs and that of existing TFs in a normalised space. As a linear plot would express no compression, it can be seen that $\mathrm{PTF}_{2.2}$ provides a small amount of compression.

3.3 Results

To evaluate how the efficiency of PTF compares with other proposed methods it has been compared with the following four state-of-the-art one-stream methods (described in more detail in Section 3.1): HDRV [MKM+04], Fraunhofer [GT11], PQ [MND13], and HLG [Bor14]. For fairness, HDRV and Fraunhofer were adapted from their original presentation for use with a 10-bit video encoder. HDRV was implemented with the luminance range 1×10^{-5} to 1×10^{4} such that the TVI curve [FPS+96] could provide a mapping from luminance to 10-bit luma. The Fraunhofer implementation uses Adaptive LogLUV [MT10] which provides mappings for a flexible number of bits.

These methods will be compared on an objective basis using the metrics presented in this section. Section 3.3.2 presents an analysis of the effect of γ on the coding error introduced by compression is provided. The results of the objective evaluation performed on the compression methods are then presented. Finally, the computational performance of PTF in contrast with PQ and look-up tables is addressed in Section 3.3.4.

3.3.1 Metrics

The following three metrics are used to provide results for the evaluation.

PSNR (Peak Signal to Noise Ratio) is one of the most widely used metrics for comparing processed image quality. To adapt the method for HDR imaging, L_{peak} was fixed at $10{,}000\,\mathrm{cd/m^2}$ and the result was taken as the mean of the channel results.

$$\mathrm{PSNR}_{\lambda} = 20 \log_{10} \left(\frac{L_{\mathrm{peak}}}{\sqrt{\mathrm{MSE}_{\lambda}}} \right) \tag{3.3}$$

puPSNR (Perceptually Uniform PSNR) was proposed as an extension to PSNR such that it is capable of handling real-world luminance levels without affecting the results for existing displays [AMS08]. The proposed metric maps the range 1×10^{-5} to $1 \times 10^{8}\,\mathrm{cd/m^2}$ in real-world luminance to values that approximate perceptually uniform values derived from a CSF. It is from the remapped luminance that the PSNR is calculated.

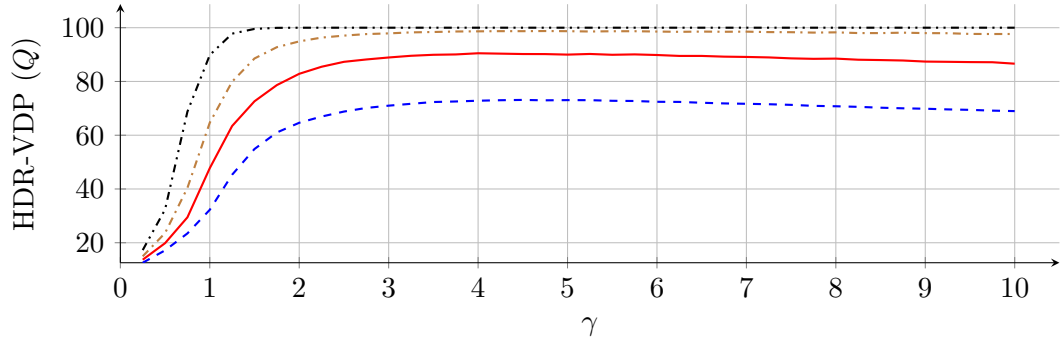
HDR-VDP-2.2.1 (HDR Visual Difference Predictor) is an objective metric based on a detailed model of human vision [MKR+11]. The metric estimates the probability at which an average human observer will detect differences between a pair of images in a psychophysical evaluation. The visual model used by this metric takes several aspects of the human visual system into account such as intra-ocular light scatter, photo-receptor spectral sensitivities and contrast sensitivity. HDR-VDP-2.2.1 has been shown to be the objective metric that correlates most highly with subjective studies [HBP+15; MDB+16a].

The metrics were calculated for every frame – except HDR-VDP-2.2.1 which was calculated for every 10th frame due to its computational expense – and averaged to produce a final figure for the sequence.

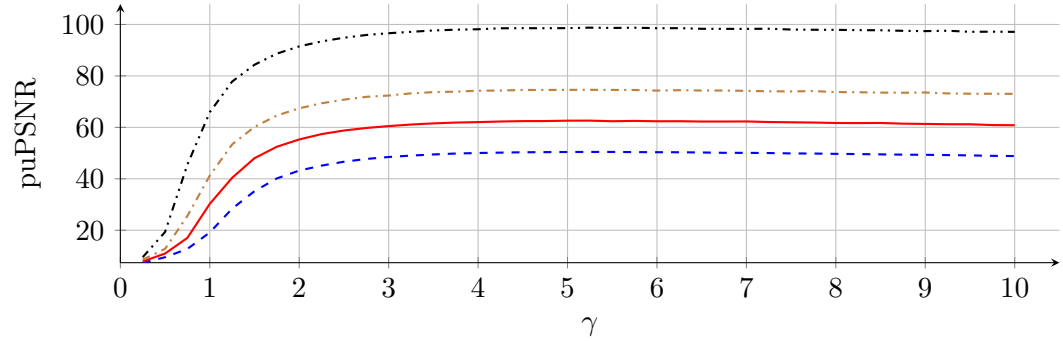
3.3.2 Analysis of Power Functions

Figures 3.4a to 3.4c show the motivation for the selection of γ by comparing the average distortion introduced by PTF over a range of γ values. These figures suggest that the different metrics favour certain γ values over others. Appendix A shows the dataset of 20 HDR images that was used for the evaluation.

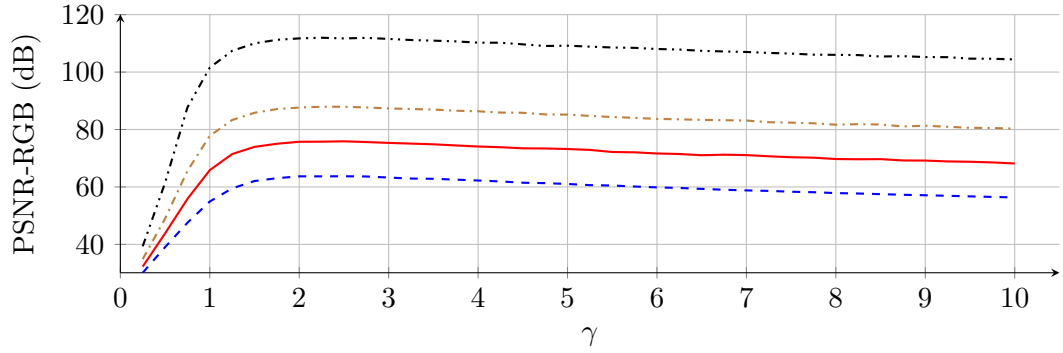
The pipeline used for this analysis is shown in Figure 3.5. After compression and colour conversion the images were not passed through the video encoder and were instead immediately decompressed to ascertain just the coding errors introduced by each γ value. The γ values used in the evaluation ranged from 0.25 to 10 and increased in steps of 0.25. The evaluation was performed at four bit-depths: 8, 10, 12 and 16. PSNR-RGB suggests that a γ of 2.2 will give the best results and, as explained in Section 3.2.3 it is the effective γ used in LDR video systems, it gives a good idea of how LDR video techniques cope with HDR data. HDR-VDP-2.2.1 Q correlate indicates that a γ of around 4 will perform best and puPSNR a γ of around 6. Figure 3.3 shows that the PQ TF proposed by Miller et al. [MND13] can be closely approximated by a γ value of 8 and hence the value was also tested. Integer values are favoured as the operations required to decode are significantly faster than non-integers, as discussed in Section 3.3.4. Based on the peaks of the graph, and similarities to the GDF and PQ



(a) HDR-VDP-2.2.1 Q correlate



(b) puPSNR



(c) PSNR-RGB

Figure 3.4: The relationship between γ and coding error for PTF created at different bit-depths across a range of metrics. The results are the average distortion introduced by PTF for the selection of HDR images in Appendix A.

(see Section 3.2.3), the four implementations of PTF chosen for testing were: $PTF_{2.2}$, PTF_4, PTF_6 and PTF_8.

Also of note in Figure 3.4 is the how the peak in quality does not shift greatly as the bit-depth is increased. This suggests that γ will not need to be changed in an environment of 12 and above bits.

3.3.3 Quality

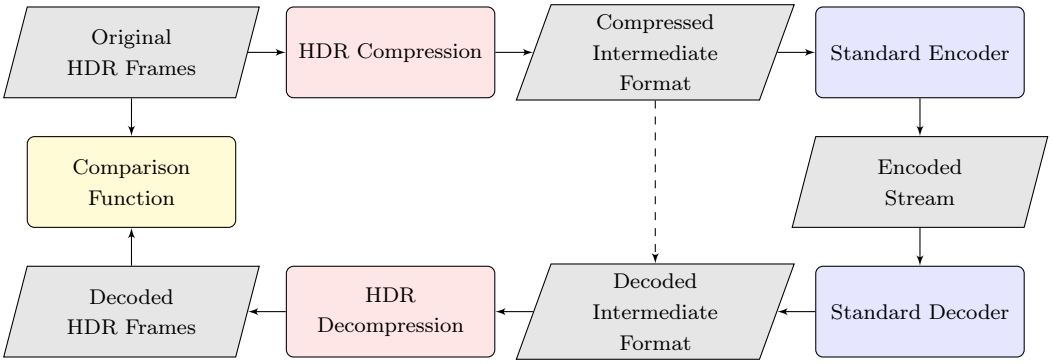


Figure 3.5: The evaluation framework used for comparing compression methods. The dashed line denotes comparison of coding errors only.

The approach used for quality comparison is outlined in Figure 3.5. For each of the compression methods the framework is executed in its entirety. The content is provided as individual HDR frames in OpenEXR format. The compression method's encoding process is run on each of the ten sequences of frames, presented in Table 3.1, to produce 10-bit files in $Y'C_BC_R$ format. These sequences were chosen as they cover a wide range of content types, such as computer graphics renderings, video captured by a SphereonVR HDR Video Camera or an ARRI Alexa. Each scene consisted of 150 frames and was encoded at 24 frames per second. The encoding was conducted with the HEVC encoder x265, due its computational efficiency, and 4:2:0 chroma subsampling with the quantisation parameters $QP \in [5, 10, 15, 20, 25, 30, 35]$. The Group Of Pictures (GOP) structure contained both bi-directional (B) and predicted (P) frames and the pattern used was (I)BBBP where the intra (I) frame period was 30 frames. The encoded bitstreams were then decoded using the HEVC Test Model (HM) [SOH+12] reference decoder, and subsequently using the individual compression method's decoding process.

3.3.3.1 Analysis

Figure 3.7 shows the results for each of the tested methods for the three quality metrics presented in Section 3.3.1. On each of the figures an increase on the Y axis indicates

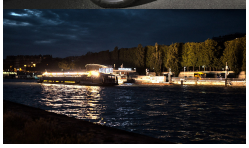
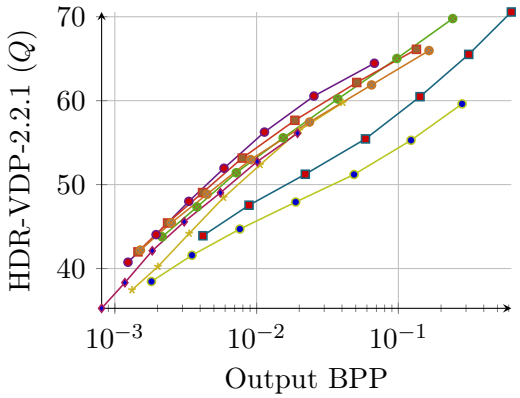
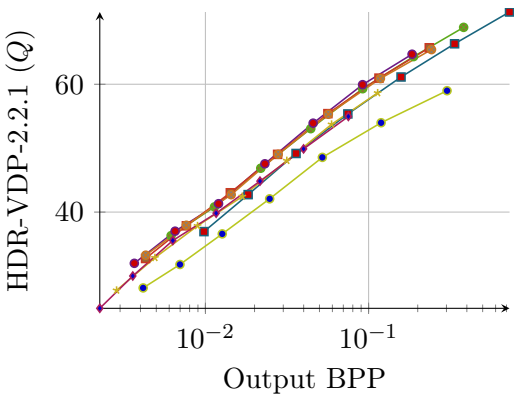
Preview	Name	Resolution	Dynamic Range (Stops)
	Welding	1920×1080	20.54
	Jaguar	1920×1080	25.30
	Seine [LLF13]	1920×1080	20.54
	Tears of Steel [Hub12]	1920×800	20.35
	Mercedes [FGE+14]	1920×1080	19.85
	Beer Festival 4 [FGE+14]	1920×1080	22.11
	Carousel Fireworks 9 [FGE+14]	1920×1080	22.38
	Bistro 3 [FGE+14]	1920×1080	22.13
	Fireplace 1 [FGE+14]	1920×1080	22.45
	Showgirl 1 [FGE+14]	1920×1080	22.73

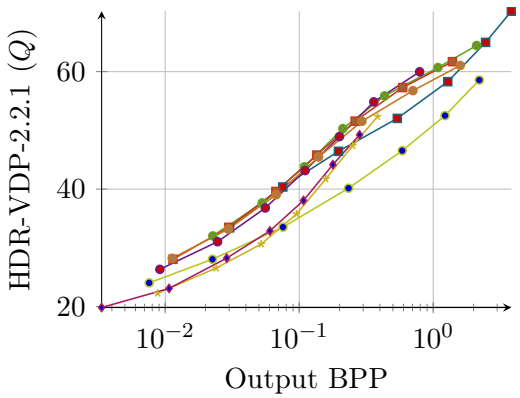
Table 3.1: The ten HDR video sequences used to evaluate the methods, showing resolution and dynamic range.



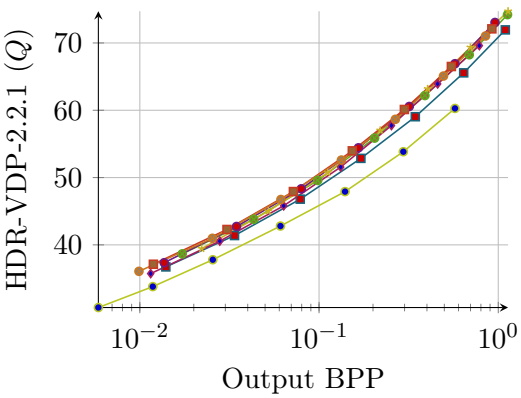
(a) Welding



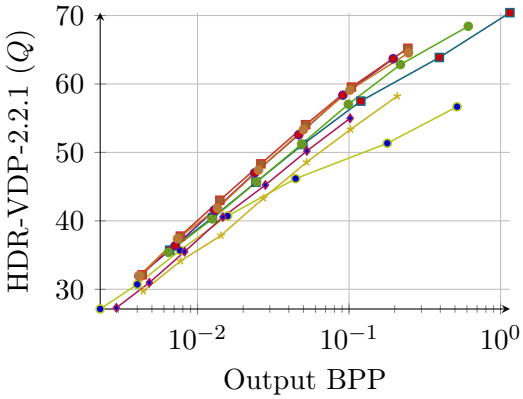
(b) Jaguar



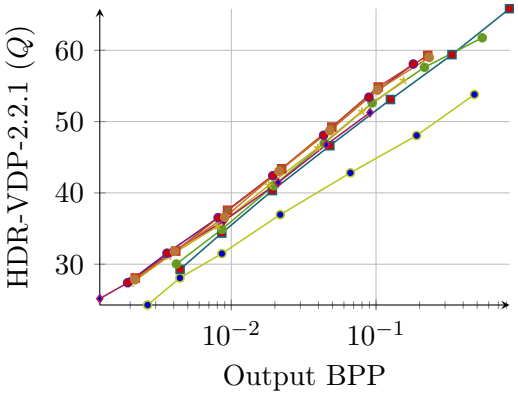
(c) Seine



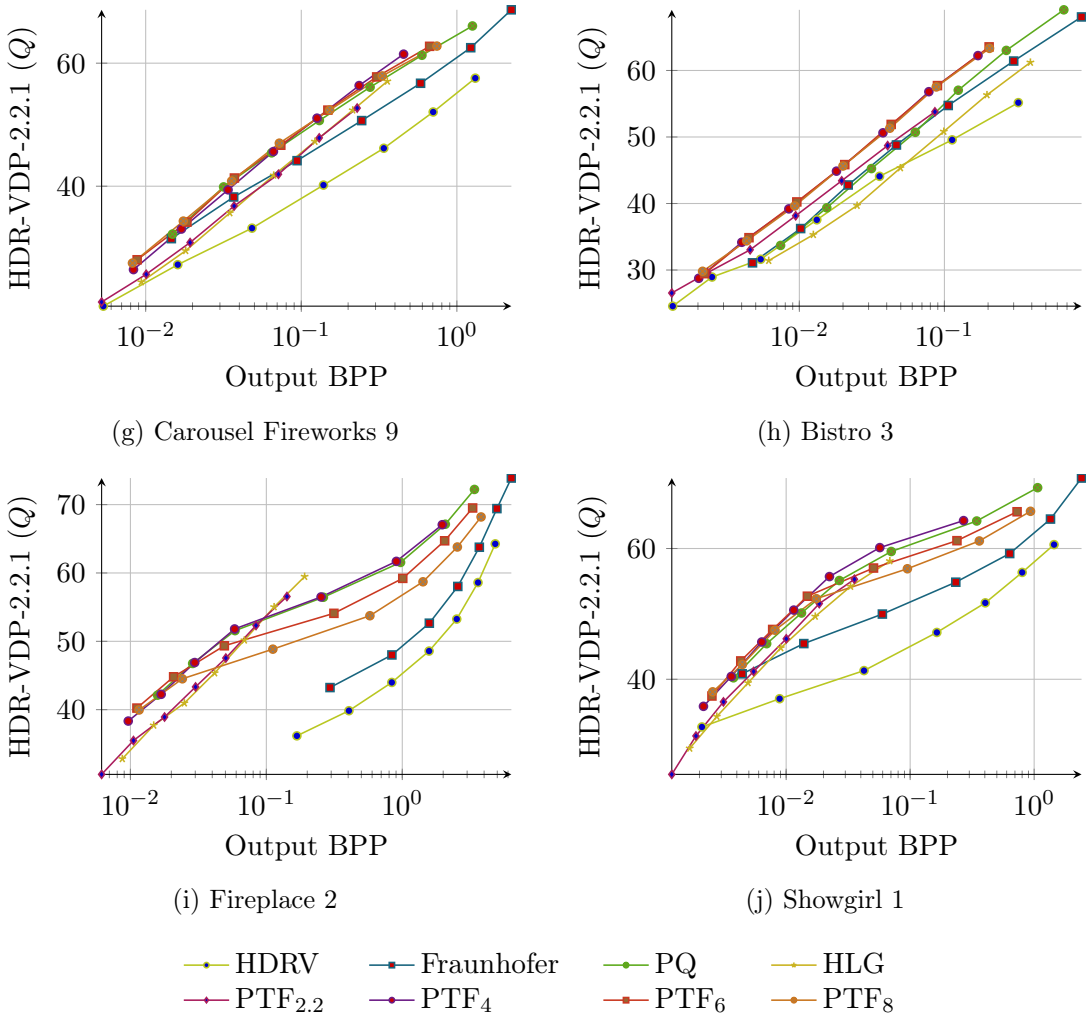
(d) Tears of Steel



(e) Mercedes



(f) Beer Festival 4



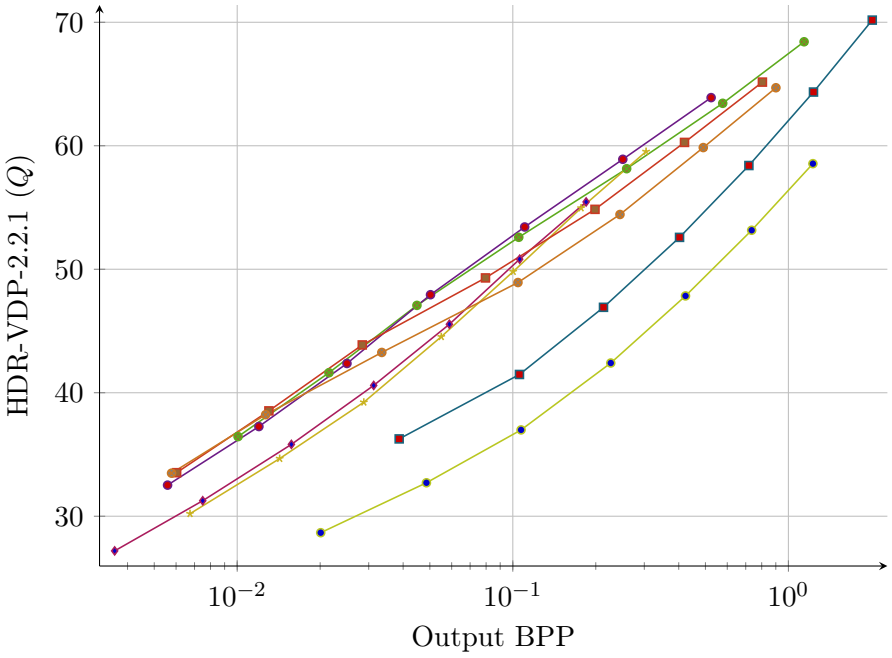
(g) Carousel Fireworks 9

(h) Bistro 3

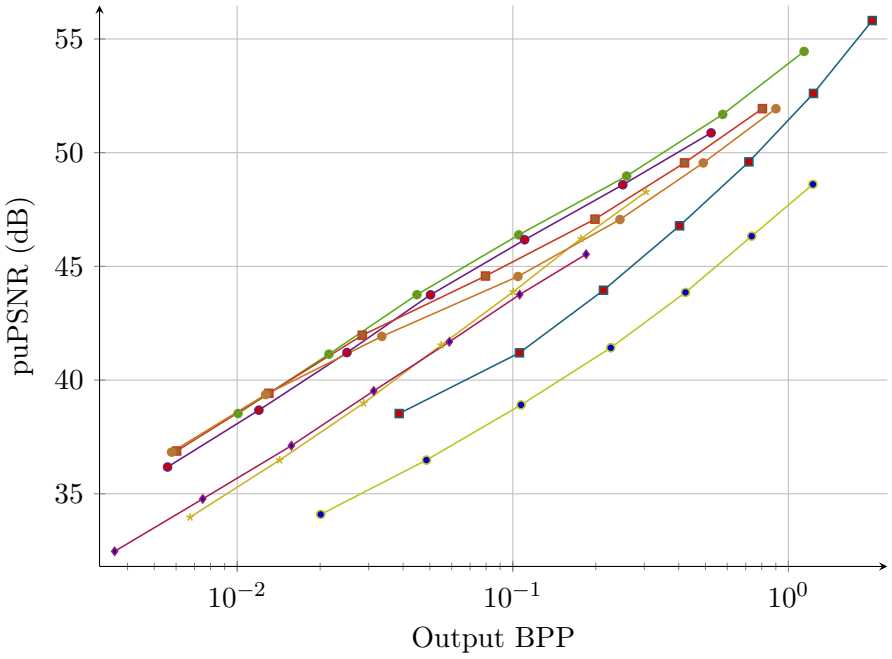
(i) Fireplace 2

(j) Showgirl 1

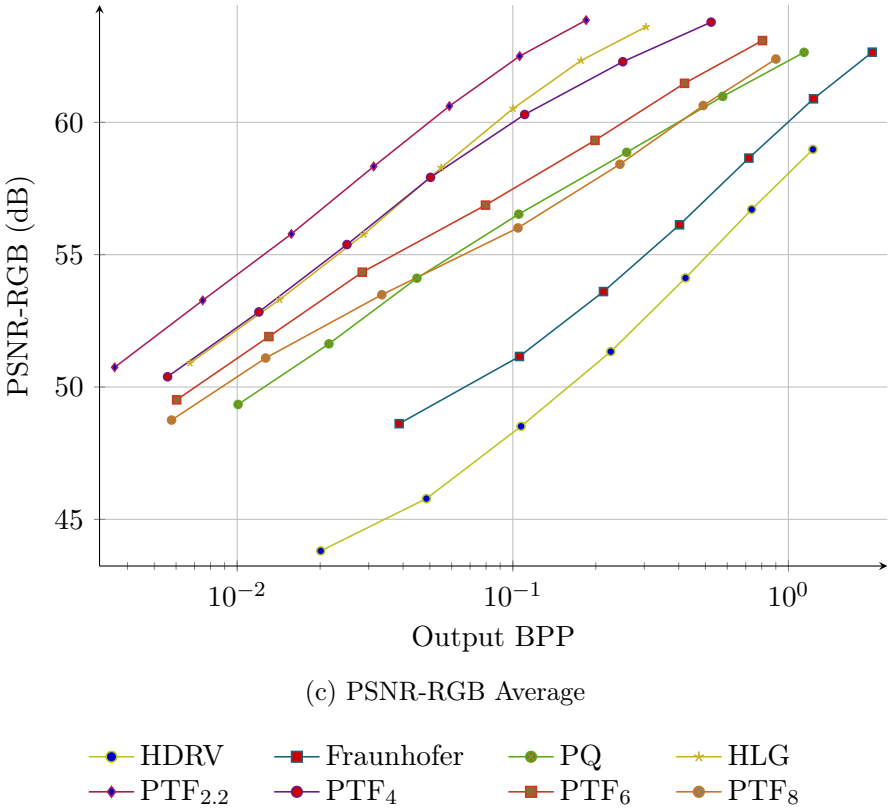
Figure 3.6: Rate distortion characteristics showing how the different HDR video compression methods perform on a variety of sequences. The rate is measured in output bits per pixel (BPP) and the distortion as a HDR-VDP-2.2.1 Q correlate. Figures are presented with a logarithmic x axis to improve clarity.



(a) HDR-VDP-2.2.1 Average



(b) puPSNR Average



(c) PSNR-RGB Average

Figure 3.7: Rate-distortion characteristics showing the results of each method averaged over the ten sequences and with three metrics, HDR-VDP-2.2.1, PSNR and puPSNR. The Output BPP is shown on a logarithmic scale to improve clarity.

improved objective quality, and an increase on the X axis indicates increase bit-rate and therefore filesize. Therefore results closest to the top-left corner are preferred. For each method at each QP, the average BPP of the encoded bitstreams across all sequences is calculated and plotted against the average quality measured. The ten HDR video sequences used to test the compression methods are shown in Table 3.1. HDR-VDP-2 results for each sequence are presented in Figure 3.6, and puPSNR and PSNR results are presented in Appendix B.

3.3.3.2 Discussion

The rate-distortion plots shown in Figure 3.7 present the trade-off between bit-rate and quality for each method. If a plotted line maintains a position above another, this indicates that improved quality can be consistently obtained from a method even with a reduction in bit-rate.

These figures show that $PTF_{2.2}$ achieves the highest average PSNR followed by HLG

	$PTF_{2.2}$	PTF_4	PTF_6	PTF_8	HDRV	Fraun.	PQ	HLG
$PTF_{2.2}$	0.00	−2.39	−1.11	0.14	11.32	5.65	−1.18	0.42
PTF_4	2.39	0.00	0.96	2.24	13.35	7.28	0.32	2.90
PTF_6	1.11	−0.96	0.00	1.30	12.52	6.74	−0.62	1.58
PTF_8	−0.14	−2.24	−1.30	0.00	11.18	5.42	−1.91	0.20
HDRV	−11.32	−13.35	−12.52	−11.18	0.00	−5.10	−13.64	−11.39
Fraunhofer	−5.65	−7.28	−6.74	−5.42	5.10	0.00	−7.95	−5.87
PQ	1.18	−0.32	0.62	1.91	13.64	7.95	0.00	1.93
HLG	−0.42	−2.90	−1.58	−0.20	11.39	5.87	−1.93	0.00

Table 3.2: Bjøntegaard delta VDP results showing the average improvement in HDR-VDP-2.2.1 Q correlate results between pairs of methods over ten sequences. Positive numbers denote a HDR-VDP-2.2.1 Q correlate improvement on average over the range of bit-rates exhibited by the row method on the left verses the column method above.

	$PTF_{2.2}$	PTF_4	PTF_6	PTF_8	HDRV	Fraun.	PQ	HLG
$PTF_{2.2}$	0.0 %	42.6 %	22.8 %	7.7 %	−80.9 %	−59.1 %	24.4 %	−5.8 %
PTF_4	−29.9 %	0.0 %	−13.3 %	−29.1 %	−86.4 %	−70.1 %	−4.8 %	−33.2 %
PTF_6	−18.6 %	15.4 %	0.0 %	−18.9 %	−84.2 %	−66.2 %	9.8 %	−22.6 %
PTF_8	−7.1 %	41.0 %	23.3 %	0.0 %	−81.6 %	−59.1 %	33.8 %	−8.2 %
HDRV	424.6 %	635.9 %	532.0 %	444.4 %	0.0 %	101.9 %	536.9 %	397.5 %
Fraun.	144.7 %	234.9 %	195.5 %	144.6 %	−50.5 %	0.0 %	219.0 %	138.2 %
PQ	−19.6 %	5.1 %	−8.9 %	−25.3 %	−84.3 %	−68.7 %	0.0 %	−25.4 %
HLG	6.2 %	49.7 %	29.2 %	8.9 %	−79.9 %	−58.0 %	34.0 %	0.0 %

Table 3.3: Bjøntegaard delta bit-rate results showing the average reduction in bit-rate achieved to maintain quality over a ten sequences. Negative numbers denote the percentage saving in bit-rate by the row method to result in a sequence with the same HDR-VDP-2.1.1 Q correlate result as the column method.

then PTF_4. The PSNR metric does not make use of a perceptual weighting on the error between the reference frame and the distorted frame. Due to this, $PTF_{2.2}$ is rated highly because the close to linear mapping it provides reduces error in the bright regions while failing to preserving detail in the dark regions. The reduced error on the relatively large values found in the bright regions therefore favour $PTF_{2.2}$ when tested with PSNR.

HDR-VDP-2.2.1 [MKR+11] and puPSNR [AMS08] use perceptual weightings that recognise that error in the dark regions is more noticeable to the HVS than the error in the bright regions. These metrics show that on average PTF_4 exhibits the least error for a given bit-rate than the other methods, although for certain sequences, such as Beer Festival 4, PTF_6 achieves the highest quality. PTF_4 preserves more data in the dark regions than $PTF_{2.2}$ but less than PTF_6 or PTF_8.

The Bjøntegaard delta metric [Bjø01] calculates the average difference in quality between pairs of methods encoding sequences at the same bit-rate. This metric will determine the average HDR-VDP-2.2.1 Q correlate gain over the range of bit-rates

achieved by PTF when compared with the other methods evaluated. From Table 3.2 it can be seen that PTF_4 gained 0.32 over PQ, 2.90 over HLG, 7.28 over Fraunhofer and 13.35 over HDRV. It can also be seen that PTF_4 gained 0.96 over PTF_6, 2.24 over PTF_8 and 2.39 over $PTF_{2.2}$. The Bjøntegaard delta bit-rate metric shows the precentage reduction in bit-rate to achieve the same objective quality. Table 3.3 shows these results.

3.3.4 Computational performance

	Time per Frame (ms)				**Speed Up (ratio)**	
	Analytic		**LUT**		**PTF'_4**	
Name	**PTF'_4**	**PQ**	**PTF'_4**	**PQ**	**PQ**	**LUT**
Welding	2.57	66.37	4.13	3.95	25.85	1.61
Jaguar	2.73	66.78	3.92	3.87	24.47	1.44
Seine	2.58	64.01	3.92	3.92	24.86	1.52
Tears of Steel	2.69	98.08	3.95	3.91	36.49	1.47
Mercedes	2.72	73.57	3.80	3.95	27.00	1.39
Beer Festival 4	2.61	65.16	3.73	3.81	24.92	1.43
Carousel Fireworks 9	2.56	65.91	3.77	3.93	25.79	1.48
Bistro 3	2.63	65.85	3.82	3.95	25.00	1.45
Fireplace 2	2.31	129.84	3.66	3.86	56.22	1.58
Showgirl 1	2.70	69.39	3.89	3.99	25.69	1.44
Average	2.61	76.50	3.86	3.91	29.63	1.48

Table 3.4: Difference in decoding time per frame between PTF'_4, $PQ_{forward}$ and their LUT equivalents across a range of sequences and averaged over five tests per sequence performed on a workstation PC. Speed up is ratio between and $PQ_{forward}$ and between PTF'_4, and PTF'_4 and the LUT implementation of PTF_4. Time per frame is presented graphically in Figure 3.8.

High performance is essential for real-world encoding and decoding. With that in mind, PTF was compared against an analytical implementation of PQ [MND13] and against look-up tables (LUTs).

Table 3.4 shows the decoding performance of PTF'_4 and PQ and their LUT equivalents, PTF'_4 LUT and PQ LUT, for the scenes presented in Table 3.1. The 1D LUTs were generated by storing the result of each transfer function for every 10-bit input value in a floating-point array. The scaling required to reconstruct the full HDR frame was also included in the table to improve performance resulting in a mapping from

Figure 3.8: Difference in decoding time in frames per second between PTF'_4, $\mathrm{PQ}_{\mathrm{forward}}$ and their LUT equivalents. Higher bars indicate faster decoding performance.

10-bit compressed RGB to full HDR floating-point. The results were produced by a single-threaded C++ implementation compiled with the Intel C++ Compiler v16.0. Only the inner loop was timed so disk read and write speeds are not taken into account. Each result was taken as the average of five tests per method on each sequence to reduce the variance associated with CPU timing. The software was compiled with the AVX2 instruction set with automatic loop-unrolling, O3 optimisations and fast floating-point calculations. The machine used to run the performance tests was an Intel Xeon E3-1245v3 running at 3.4 GHz with 16 GB of RAM running the Microsoft Windows 8.1 x86-64 operating system.

Encoding performance was also evaluated. In this case the mapping was from full HDR floating-point to 10-bit output and hence the LUT implementations could not include scaling in the table. The compression methods, sequences, resolution and sequence lengths were the same as above. PTF_4 encoding was achieved on average per frame in 4.37 ms, PQ encoding in 72.59 ms, PTF_4 LUT in 4.02 ms and PQ LUT in 4.21 ms.

Listing 1 PTF_4 Optimised Decoding C++ and associated x86-64 AVX2 Assembly

```
inline float g4decode(float in, float norm, float trough)
{
    float x2 = in * in;
    float lin = x2 * x2;
    float hdr = lin * norm;
    return hdr;
}
```

```
vmovups   ymm0, ymmword ptr [rdi+rbx*4]
vmulps                  ymm3, ymm0, ymm0
vmulps                  ymm2, ymm3, ymm3
vmulps                  ymm3, ymm2, ymm1
vmovups   ymmword ptr [rdi+rbx*4], ymm3
```

The results demonstrate that the straightforward floating-point calculations required to decode PTF_4 can outperform the floating-point calculations required to decode PQ by a factor of 29.63 times and even the indexing needed to use a look-up table by 1.48 times. The high performance of PTF'_4 is due to its compilation into only a few instructions, in this case three multiplies, that can have high performance SIMD implementations. PTF also avoids any branching, improving performance on pipelined architectures. The Intel C Compiler generated the Assembly shown on the right in Listing 1 from the C++ code shown on the left. Encoding PTF_4 can be achieved at a speed comparable to the use of LUT and greatly in excess of an analytic implementation of PQ.

3.4 Conclusion & Future Work

This chapter has introduced and evaluated a straightforward method of compressing HDR video streams. It has shown that a transfer function based on power functions

is capable of producing high-quality compressed HDR video. It has also shown that compression can be achieved using straightforward techniques which lend themselves to implementation in real-time and low-power environments. On a commodity desktop machine, PTF is able to be decoded at over 380 fps and outperforms an analytic implementation of PQ by a factor of over 29.5 and a look-up implementation by a factor of nearly 1.5. Encoding performance outperforms PQ by a factor of 16.6 and is only slightly slower than a LUT. Thanks to its straightforward nature, PTF is amiable to acceleration through the use of hardware such as FPGAs and GPUs. This chapter has demonstrated that high-quality HDR video compression does not require complex transfer functions.

The method possesses a parameter that adjusts the strength of the compression. As a result of the evaluation, and the desire to optimise the method for computational performance, the parameter (γ) was set to the integer value 4 as it represented the single value which on average gave the best objective results. However, in some sequences other values resulted in a higher objective result. It should also be noted that only a small range of values were tested, this was to gain an understanding of the space occupied by the parameter but also to present a concise evaluation. Chapter 5 will investigate this parameter space in much greater detail to ascertain whether improved compression can be gained by fine-tuning the parameter further.

Chapter 4

Effects of Ambient Illumination on Perceived Detail

High dynamic range content is applicable to a broad range of scenarios. Chapter 3 proposed an HDR video compression method that was shown to compare favourably against other state-of-the-art methods when compressing fixed content. This chapter will investigate if other variables may need to be taken in to account when designing a HDR compression system, such as: display brightness and the environment at the display.

Often the ambient environment at the display will not be known at the time of compression. In many, the ambient illumination will vary greatly over time. Some examples include an office environment where the illumination will change throughout the day, as the sun moves through the sky and clouds pass overhead; the television situated in front of a window, where, during the time most content is consumed, the sun may be low in the sky, casting rays directly on to the screen; or an outdoor advertising screen or public television where at any given moment the environment may be, for example, sunny, cloudy or raining. This chapter will demonstrate how the ambient environment in which content is viewed, and specifically the illumination incident on the display, affects the perceived detail visible to the subject.

Akyüz et al. [AFR+07] demonstrated that participants prefer a HDR image over tone-mapped or LDR image of a scene, Mukherjee et al. [MDB+16b] showed that this also holds true for HDR video. Melo et al. [MBB+15] investigated tone-mapped HDR content on an LDR display to find that the ambient illumination, and especially reflections on the screen, significantly affected the viewing experience. An experiment in which participants watched HDR video on an HDR display under different ambient illuminations showed that visual fatigue was not a particular problem [RHL+09]. There

was, however, a significant difference in the choice of brightness and contrast depending on the lighting conditions. This work conducted two experiments on a display with a peak luminance of 4,000 cd/m^2 and a black level of <0.001 cd/m^2 at seven ambient illuminations. <0.01, 0.75, 8.5, 28 and 74 lx in the first and, <0.01, 70 and 700 lx in the second. In an experiment conducted in a dark room, Daly et al. [DKS+13a] showed that for diffuse imagery the average preferred range of display was between 0.1 to 650 cd/m^2. However to satisfy 90% of the population the diffuse range should be between 0.005 to 3,000 cd/m^2 and greater than 20,000 cd/m^2 for highlights [DKS+13b].

4.1 Motivation

There have been no studies presented to date that investigate the combination of both high peak luminance displays and bright ambient environments. Peak brightnesses of up to 20,000 cd/m^2 [DKS+13a] and ambient illuminations of up to 700 lx [RHL+09] have been tested but never in combination and not in environments such as those found outdoors where the ambient illumination can greatly exceed those previously tested.

This chapter will present an experiment in which both a high peak brightness and a high ambient illumination are tested. The experiment was performed on a prototype SIM2 [SIM15] display capable of producing 10,000 cd/m^2, five times that of consumer displays, and in a range of environments up to 80,000 lx. The aim of this experiment is to measure the dynamic range that can be perceived by a participant under a range of real-world ambient illuminations and display brightnesses. The hypothesis, h_1, is that in conditions such as outdoors on a sunny day, where the illumination can exceed 80,000 lx [CL03], a higher peak luminance display will allow more information in the image to be perceived. A secondary goal is to be able to calculate the required luminance of a display for content of a known range at a given ambient illumination. This will be achieved by performing a subjective test where participants are asked to report the visibility of elements on a test chart at a range of display luminances and ambient illuminations.

Although the screen has been shown to produce a peak luminance of 10,000 cd/m^2, the measured luminance at that brightness was not precise enough to be used for the evaluation. Therefore the peak luminance was reduced to 8,000 cd/m^2 to ensure consistent luminance output for the duration of the trial.

4.2 Subjective Experiment

This chapter will investigate how peak luminance and ambient illumination affects how much detail can be perceived on an HDR display. It presents an experiment to explore

Application	**Typical ambient illumination (lx)**
Home theatre [ITU12]	∼0
Advised home environment viewing conditions [ITU12]	200
Typical office lighting	500
Inside on a sunny day	3,000
Advertising display in shopping centre	10,000
Hotel pool bar	80,000

Table 4.1: The ambient illuminance of locations for some target applications for HDR displays.

the relationship between the detail visible on a display, peak luminance and ambient illumination. It will consider environments covering a large range of lighting conditions and in particular will simulate the wide variety of locations in which HDR displays could be employed in the future. Table 4.1 shows a selection of typical ambient illuminations at potential locations for HDR displays. Should HDR displays become ubiquitous, it can be expected that they will find application in all environments, such as an advertising display in a shopping centre and the bar next to a hotel pool. Ethical approval for the experiment is included as Appendix C.

4.2.1 Design

The experiment comprised three tasks designed to determine the range visible on displays at simulated luminances. These tasks are as follows:

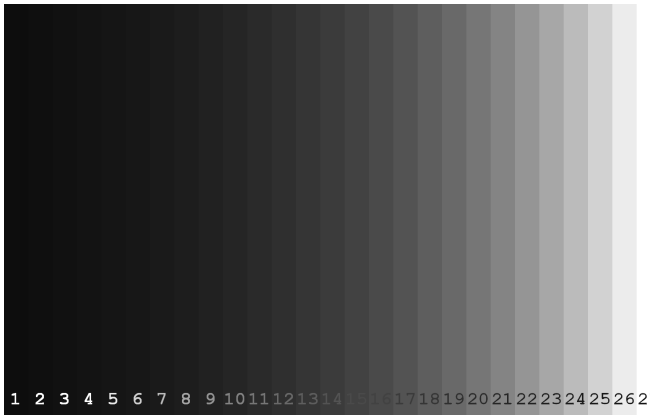
Bars: participants are asked to identify the minimum and maximum bars they can discern, shown in Figure 4.1a.

Squares: participants are asked to count the number of squares, shown in Figure 4.1b.

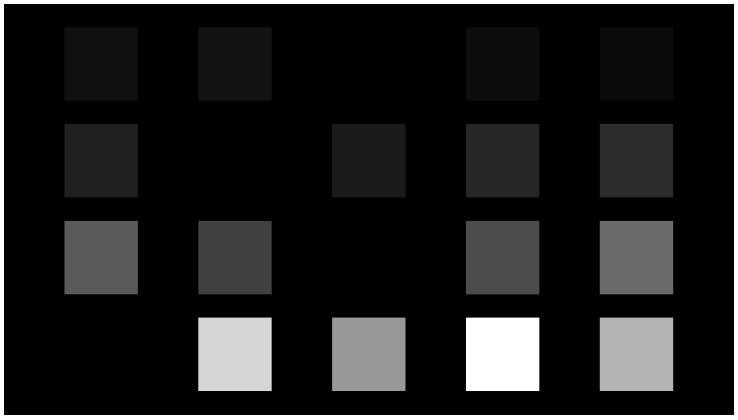
Ayres: participants are asked to give the orientation of the circles on Ayres charts [Ayr96], shown in Figure 4.1c.

These tasks have been selected as they can be conducted efficiently and provide robust results. In addition, the Ayers chart has been used previously to successfully evaluate the dynamic range of an HDR display [LCS04]. The results from each task can be used to corroborate the other two, ensuring the data is reliable. The results of the three tasks under different ambient conditions are the dependent variables of the experiment.

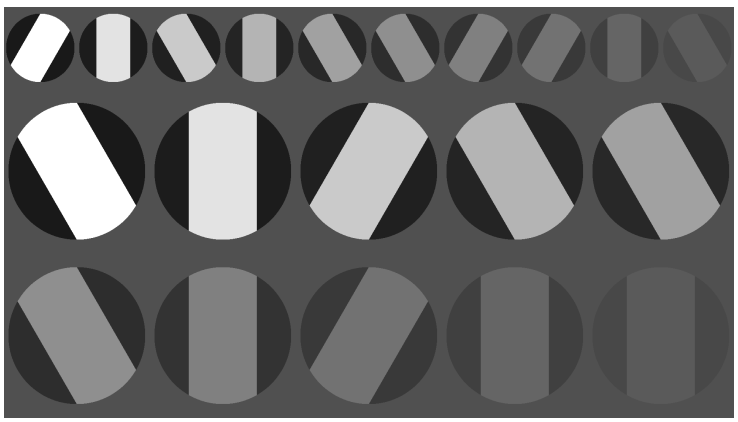
The tasks were performed under two varying conditions corresponding to the two independent variables in the experiment. The first independent variable, *brightness*, was within-participants and was the luminance of the display. Five conditions were considered:



(a) *Bars*: luminance ramp containing 30 gradations separated by 1 stop each.



(b) *Squares*: patches separated by 1.5 stops.



(c) *Ayres*: circles with a peak luminance of 20 stops above the background.

Figure 4.1: Test charts used in the subjective experiment. The charts were graded at a range of different luminances, 500, 1,000, 2,000, 4,000 and 8,000 $\mathrm{cd/m^2}$. Presented tone-mapped for print. *Squares* and *Ayres* show example patterns as the layout was randomised.

500, 1,000, 2,000, 4,000 and 8,000 cd/m^2. All luminance levels were simulated on the same display by adjusting the maximum output brightness.

The second independent variable, *ambient*, a between-participants independent variable, represents the ambient illumination in the environment. The conditions: 0, 200, 3,000 and 80,000 lx were considered. These represent no light, an indoor artificially lit working environment, an indoor sunlit (but not directly) environment, and a sunny outdoor environment respectively. The ambient illumination was be measured using a calibrated luminance meter placed in front of the display and illuminances within 1% of the target brightness were be considered acceptable. Participants viewed all *brightness* conditions for a single given *ambient* condition.

4.2.2 Materials

For each of the three dependent variables, a set of test charts were prepared and graded at a range of luminances. The luminances chosen were 8,000, 4,000, 2,000, 1,000 and 500 cd/m^2, reducing in steps of one stop. The first chart, shown in Figure 4.1b, contained 16 square luminance patches, the brightest patch shown at the peak luminance of the display with the intensity of each successive patch reduced by 1.25 stops to cover a 20 stop range. The patches were displayed on four rows with four patches on each row along with a single black patch; the order of the patches on each row was randomised so as not to allow the prediction of the location of a patch. The goal of the task was to count the number of visible patches.

The second chart, shown in Figure 4.1c, contained 20 Ayres circles with varying contrast on a grey background [Ayr96], the first row contained 10 smaller circles while the second and third rows contained 5 larger circles. Each was comprised of a dark circle with a light stripe through the middle. Each circle was presented at one of three possible orientations, $-30°$, $0°$, $30°$, chosen randomly. The circles were displayed over a range of contrasts, the strip in the highest contrast circle was chosen to be the peak luminance of the chart and was 20 stops lighter than the grey background, the dark region was chosen to be 20 stops darker than the background. Each successive circle has a lightness and darkness 2 stops less than the previous giving 20 stops above and beneath the grey background. The goal of the task was to correctly report the orientation of the circle.

The third chart, shown in Figure 4.1a, contained 30 full-height bars descending in 1 stop steps from 8,000 cd/m^2. The chart was clamped at the peak luminance required for the grading. Each bar was annotated with a number and the goal of the task was to write down the number corresponding to the first and last perceptually differentiable bar allowing the calculation of the number of stops visible.

The SIM2 display was used at five different luminance levels: 500, 1,000, 2,000,

4,000 and 8,000 cd/m^2 and is capable of a black level of >0.05 cd/m^2.

4.2.3 Environment

This display was positioned outdoors and the experiments were conducted at times selected from Table 4.2 so as to most closely match the target illuminances shown in Table 4.1. The actual conditions tested were 0, 200, 3,000 and 80,000 lx. To maintain the validity of the results obtained from the experiments, the two first scenarios (0 and 200 cd/m^2) were conducted in an experimental room in which all the environmental variables were controlled. The remaining conditions were conducted in an outdoor environment with the display in the shadow for the 3,000 cd/m^2 condition and in the sun for the 80,000 cd/m^2 condition. Care was taken to ensure that the ambient light level did not vary during the time the experiment was conducted.

Measured ambient illumination (lux)	**Time of day**
0	10pm
500	9pm
5,000	9am
5,000	6pm
80,000	12 noon

Table 4.2: Measured ambient illuminations in the testing environment.

4.2.4 Participants

The sample consisted of 40 participants (27 male and 13 female) with ages between 18 and 49 years old ($\mu = 23.75, \sigma = 7.0$) who volunteered from a local university. The participants were divided into eight groups of five each. This number of participants was chosen to facilitate an efficient experiment while ensuring that they all were within the required viewing angle for the display. Participants were chosen based on a non-probabilistic sampling technique as they were chosen based on their relative ease of access. Due the visual nature of the experiment, only participants with normal or corrected to normal vision where chosen. This criteria was based on assessment by an optician as to whether the participant required vision correction. If so, the participant was required to use correction during the experiment. If the participant did not have correction available they were not allowed to participate however this was reliant on the honesty of the participant to report whether they needed correction.

4.2.5 Procedure

Before starting the experiment, the participants were informed that they would be required to perform three tasks: count the number of squares visible on the display;

identify the direction of the luminous strip on each circle; and report the numbers corresponding to where they could no longer identify the limits between consecutive bars. They were then introduced to the environment in which their conditions will took place and asked to take a seat, ensuring that they can see the entirety of the display. They were also be asked to ensure their vision was corrected to normal if required. Each participant was handed a sheet of paper on a clipboard and a pen in order to report their results. Then a brief training run performed so that the reporting procedure clear and they understood what was required for each of the tasks. The luminance of the display and illuminance of the environment were noted at this point. Software then displayed a permutation of the images for the experiment on the display and the participants asked to mark down what they can perceive. When all the images have been displayed the luminance and illuminance were be measured again to ensure they remain consistent.

4.2.6 Results

Descriptive results for each of the three dependent variables are shown in Tables 4.3a to 4.3c. Figure 4.2 shows the results for all the means. Figure 4.3 shows the results of the experiments converted into stops. Figure 4.4 shows the results in stops averaged across the three dependent variables.

The data captured for *bars* comprises two values: a *minimum* and a *maximum*. For each capture the results are reported as the *minimum* subtracted from the *maximum* as this represents the dynamic range observed by the participant. For *squares*, the number of visible squares are reported. Finally, for *Ayres*, the number of successful choices is presented. No outliers were removed from the results.

4.2.6.1 *Bars* results

For the dependent variable *bars*, a 4 (*ambient*) $\times$ 5 (*brightness*) repeated measures factorial ANOVA was conducted.

For the main effect of *brightness* Maulchy's test of sphericity was violated ($p < 0.05$) and Greenhouse-Geisser corrections were applied. The main effect was significant $F(2.32, 83.46) = 48.412$, $p < 0.01$. Pairwise comparisons with Bonferroni corrections demonstrated significant differences amongst all the conditions at $p < 0.01$ for all conditions. However, 4,000 $\mathrm{cd/m^2}$ against 8,000 $\mathrm{cd/m^2}$ was significant at $p < 0.05$.

For the main effect of *ambient*, a significant difference was also noted $F(3, 36) = 4117.65$, $p < 0.01$. Pairwise comparisons with Bonferroni corrections found significant differences for all conditions at $p < 0.01$ except for the cases of 0 and 200 $\mathrm{cd/m^2}$ for which no difference was found $p = 0.90$.

		Illuminance (lx)					
		0	200	3,000	80,000		
Luminance (cd/m^2)	500	15.00	14.70	12.50	8.70	12.725	
	1,000	16.50	16.00	13.00	9.10	13.650	
	2,000	16.90	17.00	13.90	10.10	14.475	$\boldsymbol{\mu_b}$
	4,000	17.00	17.00	16.20	11.60	15.600	
	8,000	18.30	18.80	16.20	12.20	16.375	
		16.74	16.82	14.36	10.34		
		$\boldsymbol{\mu_a}$					

(a) *Bars*

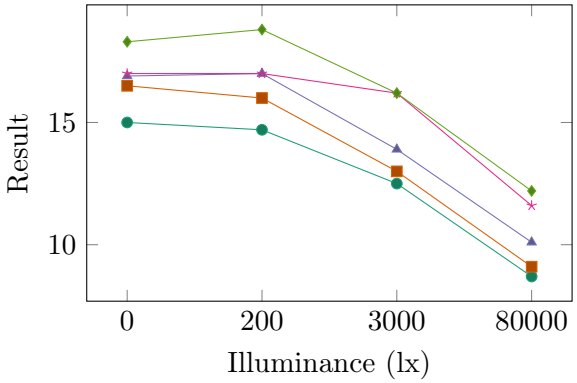
		Illuminance (lx)					
		0	200	3,000	80,000		
Luminance (cd/m^2)	500	11.50	10.60	8.90	7.70	9.675	
	1,000	12.10	11.50	9.50	7.80	10.225	
	2,000	12.00	12.00	10.30	8.20	10.625	$\boldsymbol{\mu_b}$
	4,000	12.10	12.00	11.50	9.50	11.275	
	8,000	13.20	12.60	11.70	10.50	12.000	
		12.18	11.74	10.38	8.75		
		$\boldsymbol{\mu_a}$					

(b) *Squares*

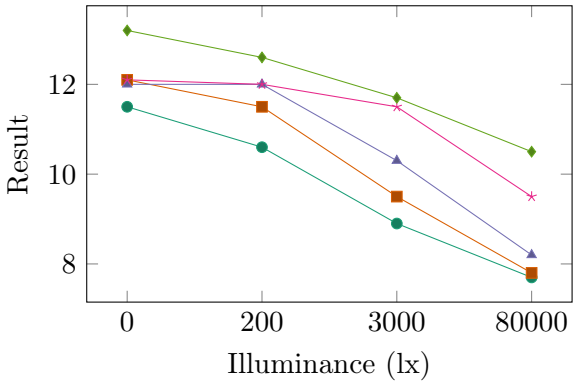
		Illuminance (lx)					
		0	200	3,000	80,000		
Luminance (cd/m^2)	500	17.00	16.00	12.70	9.90	13.925	
	1,000	18.30	16.30	14.40	11.00	15.000	
	2,000	18.60	18.00	15.10	11.80	15.875	$\boldsymbol{\mu_b}$
	4,000	19.40	18.00	15.60	12.90	16.475	
	8,000	20.00	19.20	16.70	13.70	17.400	
		18.66	17.50	14.90	11.86		
		$\boldsymbol{\mu_a}$					

(c) *Ayres*

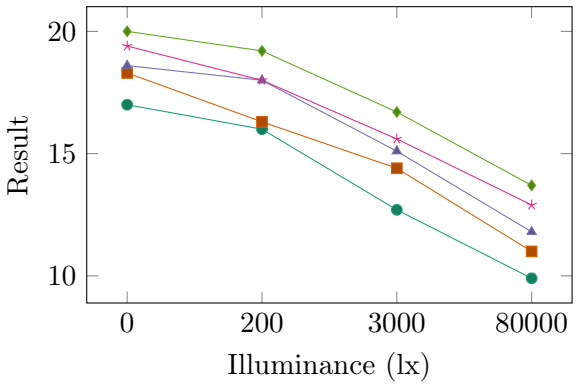
Table 4.3: Mean scores for the three dependent variables across all *brightness* and *ambient* levels. Mean correct answers are shown.



(a) *Bars* results



(b) *Squares* results



(c) *Ayres* results

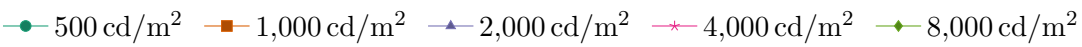
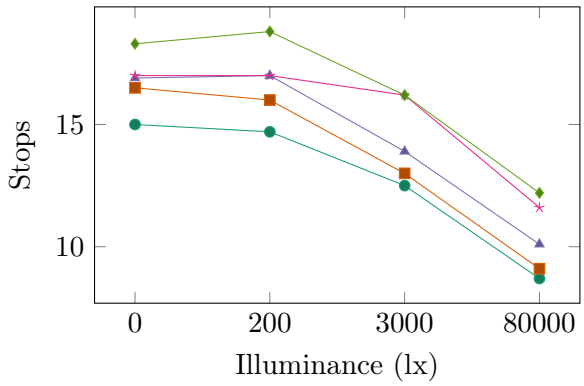
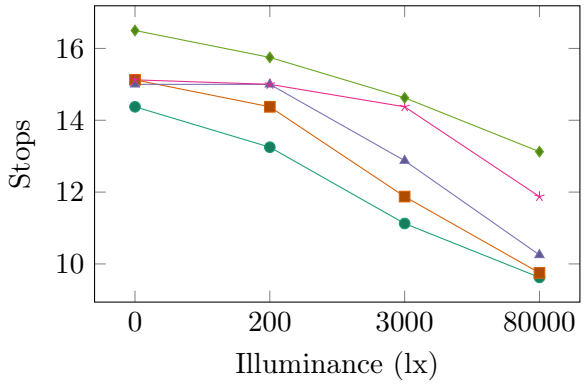


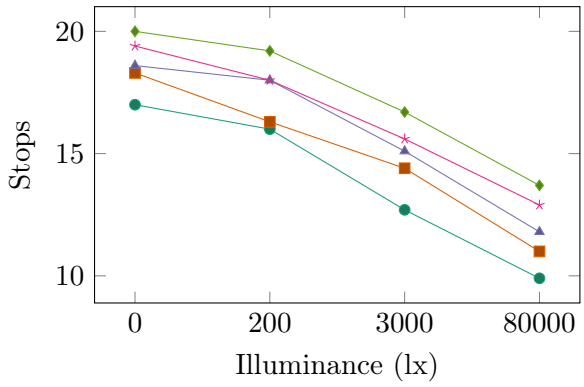
Figure 4.2: Mean scores for the three dependent variables.



(a) *Bars* results



(b) *Squares* results



(c) *Ayres* results

Figure 4.3: Mean results in stops for the three dependent variables.

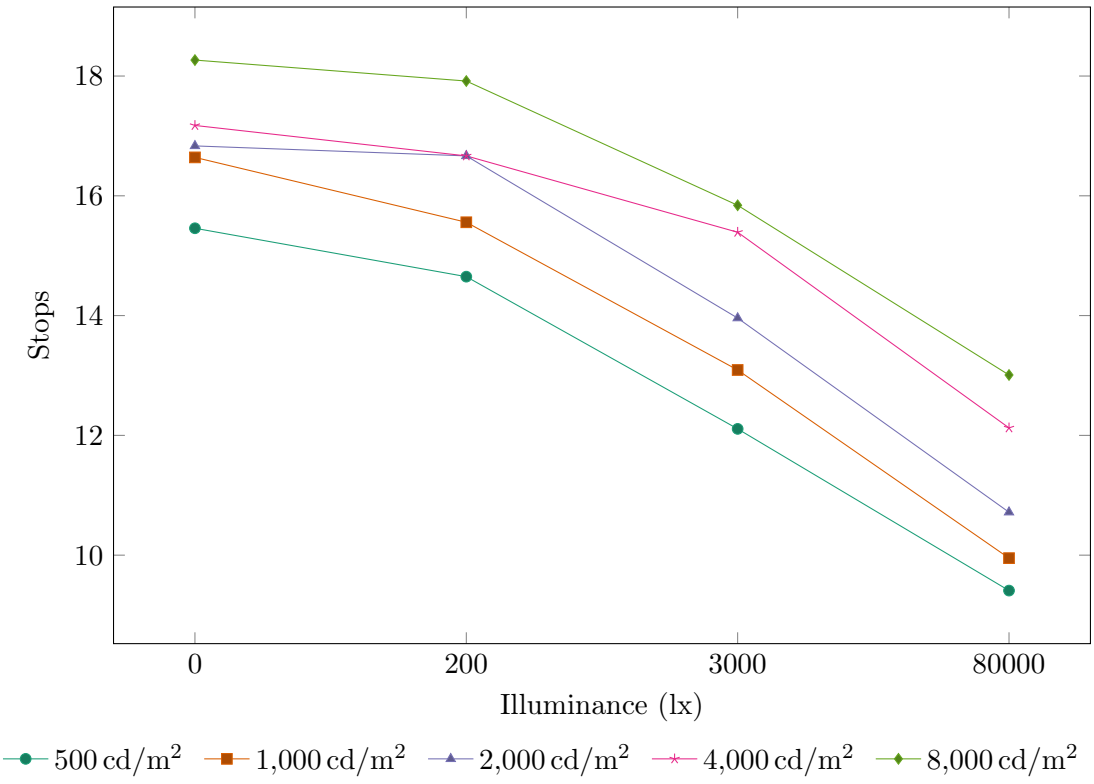


Figure 4.4: Mean results in stops for all dependent variables.

4.2.6.2 *Squares* results

The dependent variable *squares* was also analysed using a 4 (*ambient*) × 5 (*brightness*) repeated measures factorial ANOVA.

The main effect of *brightness* was found to be significant with Greenhouse-Geiser corrections as Maulchy's test for sphericity was violated, $p < 0.05$, $F(2.65, 95.56) = 63.915$, $p < 0.01$. Pairwise comparisons found significant differences across all *brightness* conditions.

The main effect of *ambient* was also found to be significant $F(3, 36) = 16155.01$, $p < 0.01$. Pairwise comparisons with Bonferroni corrections found significant differences across all conditions except 0 and 200 $\mathrm{cd/m^2}$ replicating the results seen for *bars*.

4.2.6.3 *Ayres* results

As with the other dependent variables, *Ayres* was analysed using 4 (*ambient*) × 5 (*brightness*) repeated measures factorial ANOVA. For the main effect of *brightness*, Greinhouse-Geiser corrections were applied as Maulchy's test of sphericity was violated, $p < 0.05$, and a significant effect was found $F(2.86, 103.07) = 138.47$, $p < 0.01$. Pairwise

comparisons with Bonferroni corrections were again found to be significant across all conditions.

The main effect of *ambient* was found to be significant at $F(3, 36) = 189.63$, $p < 0.01$. Pairwise comparisons with Bonferroni corrections found significant differences across all conditions.

4.2.7 Analysis

The initial hypothesis, h_1, was that in bright ambient illuminations, a higher peak brightness will allow more information from the image to be perceived and the results show that this is indeed the case. From Figure 4.4, the display luminance necessary to perceive a certain range at a given ambient illumination can be estimated, although the results for the darker conditions may depend more on the black level obtained by the display rather than the peak luminance. Around 15.5 stops can be perceived on a 500 cd/m^2 display in a 200 lx environment, whereas that increases to over 18 stops on an 8,000 cd/m^2 display. In an 80,000 lx environment, the entire range of content containing 10 stops can be perceived correctly on a 1,000 cd/m^2 display, but if the content contains 14 stops an 8,000 cd/m^2 display is needed.

4.2.8 Discussion

The results demonstrate the advantage of utilising a display with an 8,000 cd/m^2 peak luminance. Participants scores for this luminance significantly outperformed the other conditions across all the three dependent variables. It is also clear that the participants performance at each of the tasks increases as the maximum luminance of the display increases. This is to be expected and confirms the hypothesis h_1.

Another point of note is that as shown in Table 4.3b, the number of squares that can be seen on a 1,000 cd/m^2 display at 200 lx (11.5) is the same as can be seen on a 4,000 cd/m^2 display at 3,000 lx. These results provide an insight into the peak luminance required to match the dynamic range of a scene with the ambient illumination. A system which can automatically adjust to these criteria will ensure the optimal HDR viewing experience while minimising the power required as there is no need need to produce a peak luminance higher than the determined value.

The ambient illuminances chosen for this experiment are representative of those in the environments where HDR content may be viewed in the future. It was important to include a sunny environment to show it is possible, with the right peak luminance, to appreciate HDR content outside of a dark environment. Significant differences were found between ambient illuminances confirming the hypothesis that the peak luminance required to display HDR content is related to ambient illumination.

4.3 Conclusions & Future Work

As HDR content becomes more widely available, the visual enhancements it provides are likely to be desired in a wide variety of locations with a wide range of ambient environments. This evaluation has shown that for a display to provide HDR many of these environments a high peak luminance is necessary.

Furthermore, the information gained from this evaluation can be used to calculate the necessary peak luminance for the content. This implies that for optimal viewing the luminance of content is no longer fixed but should be adjusted, boosting it if the range would be lost or reducing it to save power consumption if possible. This then impacts compression as on-the-fly adjustments should be made to content to suit the environment at the display, requiring real-time compression. Chapter 5 will investigate the effects of luminance on compression quality, specifically focussing on whether a fixed transfer function is optimal for content at a range of luminances or whether an adaptive transfer function can improve quality.

Various models of the human visual system have been proposed. Further research should investigate how well the results presented in this chapter fit those models.

Chapter 5

Adapting Compression to Content and Display

Chapter 3 showed how power functions, in the form of PTF, provide high quality video compression. A selection of γ values were evaluated for use with PTF based on their quality when compressing footage. The values were chosen by looking at the mean quality across a range of images and choosing the values that correspond to the highest quality[1]. From this experiment it was found that PTF_4 provided the highest objective quality however, $PTF_{2.2}$, PTF_6 and PTF_8 were also tested to confirm the hypothesis in a more robust setting. Although PTF_4 provided the highest mean objective quality, some sequences showed that alternative γ values for PTF resulted in higher quality.

In Chapter 4 the effects of ambient illumination on perceived detail were studied. Specifically, they showed that for dynamic range to be visible in bright ambient environments, such as those found outdoors, display brightnesses of $8{,}000\,\mathrm{cd/m^2}$ are needed to fully utilise the perceivable dynamic range of the HVS with minimum adaption. However, in a dark or dimly-lit room a display with a peak brightness of $1{,}000\,\mathrm{cd/m^2}$ is capable of displaying the required range. This demonstrates that both the luminance of the display and the ambient environment at the display have an effect on what is visible to the observer and that content needs to be adjusted accordingly.

This chapter focuses on whether compression quality can be improved by adapting the PTF compression method to the content to be compressed, and to the screen for display. It shows that the optimal constant for γ changes as peak display brightness approaches $10{,}000\,\mathrm{cd/m^2}$ and how a different value is needed for optimal compression of scene-referred data. It demonstrates how the ideal value of γ varies depending on the content and how peak compression performance can only be achieved by adapting

[1] See Chapter 3 for more discussion.

the value of γ to the content. It proposes a framework for developing and evaluating adaptive compression methods and implements a new compression method, Adaptive PTF (PTF_a), around the framework. It also shows how the state-of-the-art compression methods cope with brighter luminance displays and scene-referred data and compare PTF_a with those methods. This chapter answers the following research question:

> *Can compression performance be improved by adapting the compression method to the content, the ambient environment and the luminance of the display?*

5.1 Related Work

Most compression methods, including the PQ [MND13] and HLG [Bor14] methods standardised by the International Telecommunications Union (ITU) in BT.2100 [ITU16], use a fixed transfer function. With a fixed shape transfer function, the precision with which each luminance can be represented remains constant, even at the luminance of the content changes throughout the sequence. PQ goes a step further and uses a fixed mapping between the encoded values and the luminance emitted by the screen.

Optimisation has been used in the context of compression for adapting tone-mapping functions for backwards-compatible HDR compression. Mai et al. proposed a backwards-compatible two-stream compression method comprising an LDR layer and a residual layer [MMM+11]. The residual layer is used to reconstruct the HDR frames after the LDR layer has been inverse tone-mapped. The proposed method optimises a piecewise-linear tone-curve in order to provide the most efficient tone-mapped representation of the HDR by minimising the mean squared error (MSE) between reference and distorted HDR frames. The authors then go on to propose a closed-form solution to their method by building and applying a distortion model however they do not consider the associated rate cost of the LDR layer [GRG+17]. Gommelet et al. address this problem in their work by adding a constraint to the perceptual quality of the SDR layer with the intention of preserving the artistic intent of the content [GRG+17]. They propose a new distortion model based on image gradient with the aim of better approximating the rate-distortion performance of video encoders. However, both of these methods are primarily suited to backwards-compatible HDR video compression.

Adaptive LogLUV compression proposed by Motra and Thoma and made temporally coherent by Garbas and Thoma will adjust to the luminance range required however relies on a fixed, logarithmic transfer function [MT10; GT11].

Lu et al. proposed an adaptive extension to PQ that re-allocates code words in the PQ mapping to increase quality [LPY+16b]. Operating in the IPT colour-space, the

signal range is divided up in to a number of bins with the codewords redistributed between the bins based on their impact to HDR-VDP-2 quality [MKR+11]. This has the effect of reshaping the PQ signal used for compression. This distribution of code words amongst the bins is mapped first to a LUT and then to the inverse of that LUT for decompression, the inverse LUT is then approximated by a piecewise polynomial to be passed as metadata to the decoder. The piecewise polynomial is then evaluated to generate the actual inverse LUT from which the final forward LUT is calculated. For optimisation the reshaping can be only recalculated at scene changes, this has the added benefit of reducing the amount of metadata required.

Adaptive γ parameters have been employed in the field of inverse tone-mapping. Bist et al. proposed an expansion operator (EO) which applied a power function to the LDR image to increase contrast while increasing the range to that required for HDR imagery [BCM+17]. The γ value was derived from a user study which asked the participants to select the amount of γ they preferred from the range 1.0–2.2. The results are modelled using the equation:

$$\gamma = 1 + \log_{10}\left(\frac{1}{\overline{L^*}}\right), \tag{5.1}$$

where $\overline{L^*} = \frac{L^*_{median}}{100}$ and L^*_{median} is the median of the lightness image channel clipped to the interval $[0.05 : 0.95]$ to remove outliers. The resulting tone expansion model is expressed as: $Y_{HDR} = L_{max} \times Y^{\gamma}_{SDR}$, where Y_{HDR} is the Y component of the HDR image in the XYZ colourspace; Y_{SDR} is the Y component of the SDR image in XYZ; and L_{max} is the peak luminance of the target HDR display.

5.2 Motivation

As discussed in Chapter 4, the ambient environment at the display plays a crucial role in what is perceived by an observer. To maximise the visible range of the content, its luminance may have to be adjusted. If the environment is known at the time of compression, that can be achieved as a pre-processing step before encoding. However, if, as is often the case, the environment is not known, or is variable as to be expected in a broadcast environment, the way to maximise visible range is to broadcast a scene-referred representation of the content and adjust the displayed luminance at the display. In both cases however, adjustments to the content may be required.

Figures 5.1 and 5.2 demonstrate how the selection of γ affects objective quality on an image-by-image basis. They show that significant gains in quality can be achieved by adjusting the value of γ depending on the image to be compressed. Content, it

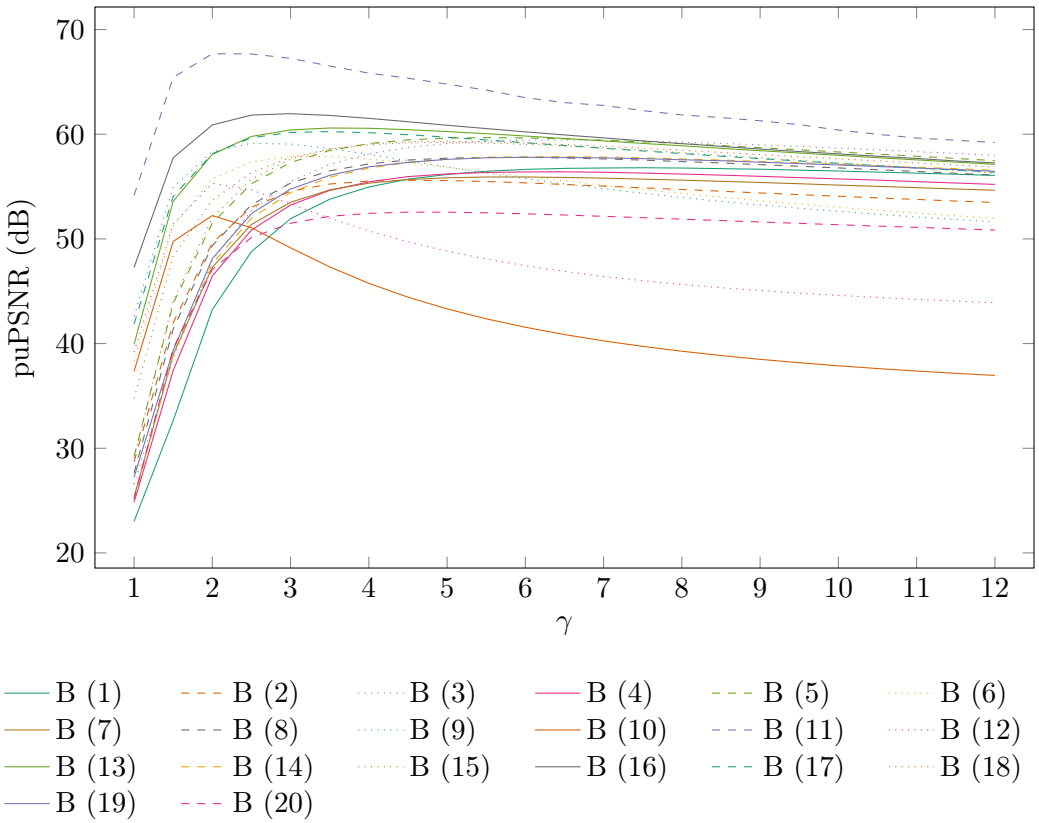
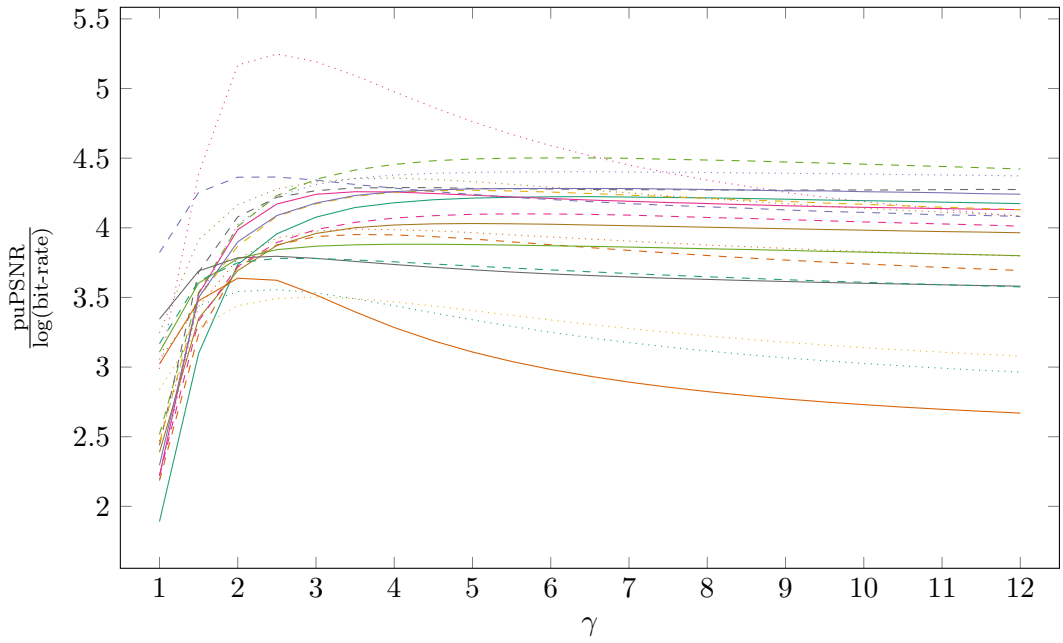
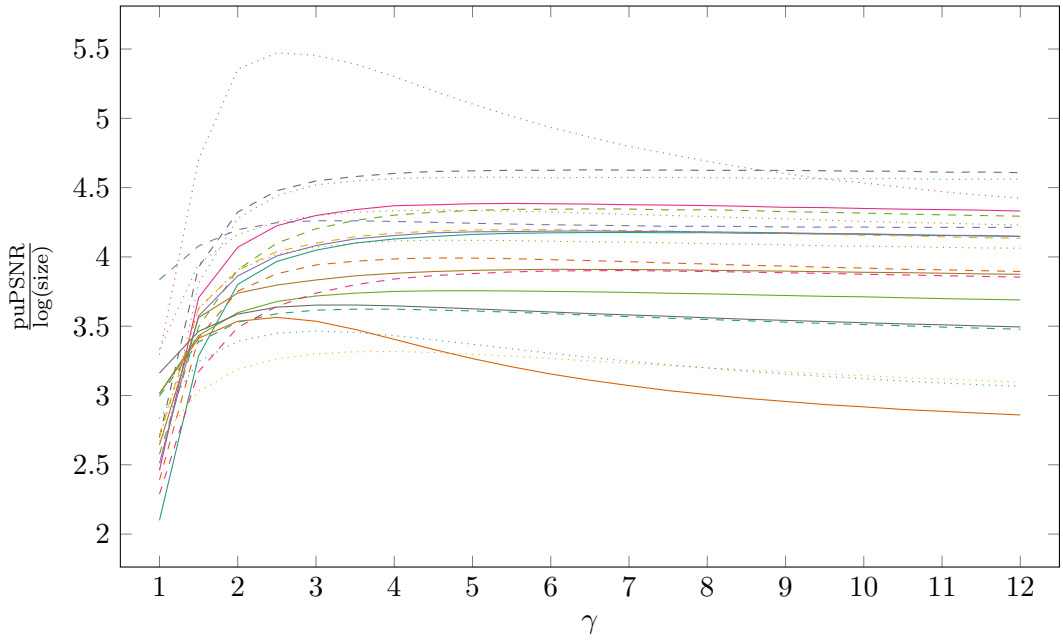


Figure 5.1: Coding-only quality in puPSNR as a function of γ for individual images when compressed with PTF. The 20 images shown in Appendix A were compressed with PTF_γ to $\text{Y}'\text{C}_\text{B}\text{C}_\text{R}$ 4:2:0 to get coding-only errors.

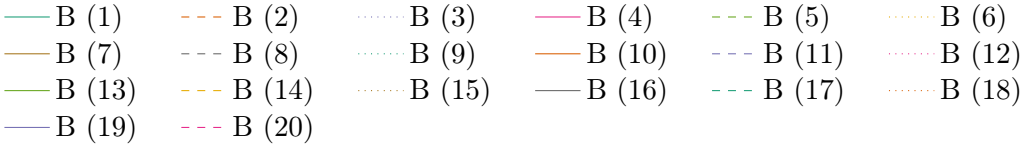
appears, plays a significant factor in the shape of the transfer function required for optimal compression.

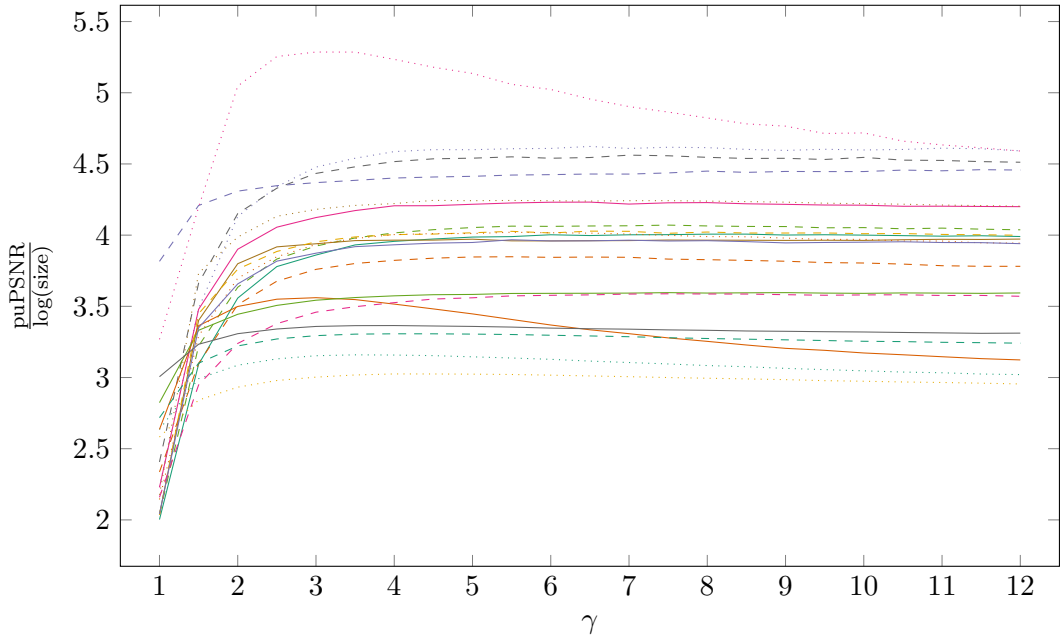


(a) x265 QP0



(b) x265 QP15





(c) x265 QP30

Figure 5.2: Compression quality in puPSNR as a function of γ for individual images when compressed with PTF and x265 at a range of QPs. The 20 images shown in Appendix A were encoded in $Y'C_BC_R$ 4:2:0 and compressed with 10-bit x265 in at 3 QP settings: (a) QP0, (b) QP15, and (c) QP30. Quality was normalised with the bit-rate to remove effects of unequal allocation of bits at the same QP value. The logarithm of the bit-rate was chosen as bit-rate grows exponentially with quality; this is demonstrated in Figure 3.7 where approximately linear plots are produced when bit-rate is plotted on a logarithmic x axis.

Quality parameter (QP) controls the quality of the encoded video by means of bit-rate control, with 0 being the highest quality (and therefore bit-rate) and 51 being the lowest. This is achieved by varying the amount of quantisation performed during the forward transform of the discrete cosine transform (DCT).

Quality Parameter	**Mean Peak** $\frac{puPSNR}{\log(size)}$	**Mean Peak** γ
0	4.1213	4.03
15	4.0937	4.85
30	3.9461	6.25

Table 5.1: Peak $\frac{puPSNR}{\log(size)}$ results with their associated γ values taken from Figure 5.2. This table shows how mean peak γ increases as the QP parameter of the encoder is increased.

Also of note is that the variation in quality is reduced as the QP is increased and the bit-rate reduced. Figure 5.3 provides an general overview of how γ affects normalised quality across different encoding QPs. Table 5.1 presents the same ef-

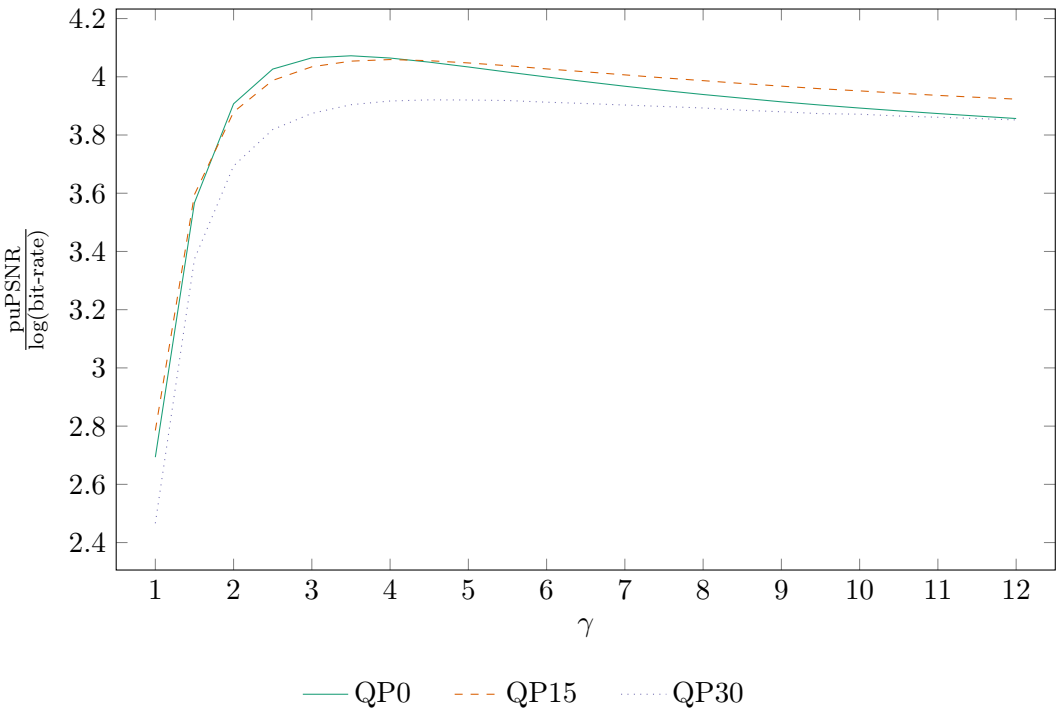


Figure 5.3: Mean compression quality as a function of γ when averaged across a selection of images compressed with PTF and x265 at a range of QPs. The 20 images shown in Appendix A were encoded in $Y'C_BC_R$ 4:2:0 and compressed with 10-bit x265 in at 3 QP settings: QP0, QP15, and QP30.

fect in numbers and shows how the mean peak γ value increases as the QP of the encoder is increased. This demonstrates how the optimal γ value for base-band quantisation is dependent on settings chosen for the encoding.

When these effects are put together, it can be seen that numerous factors, such as the ambient environment at the display and the luminance of the screen, affect how the content needs to be presented. These factors, along with features of the content and the compression such as QP, then affect the transfer function that needs to be applied to maximum compression quality. A system is therefore needed to adapt the transfer function to achieve the highest possible compression quality given the set of environmental factors.

5.3 Adaptive Compression Framework

A broad overview of how adaptive compression is performed in this thesis is presented in Figure 5.4. It shows how content, encoding parameters, and external features such as the ambient environment and luminance of the display, are taken into account by

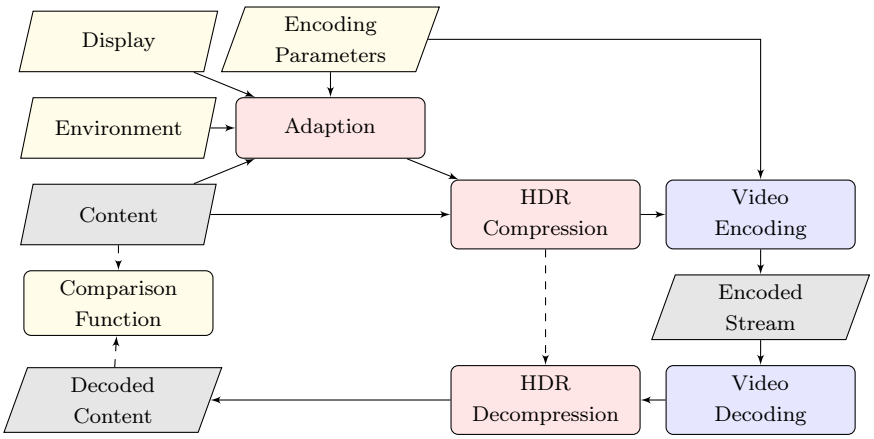


Figure 5.4: A generic framework for adaptive compression. This framework takes information from the content, environment and the display to provide parameters to the HDR compression method through an adaption method. The final decoded content can be tested with a comparison function such as puPSNR or HDR-VDP-2. Dashed lines show optional processes.

an adaption method which provides the metadata required to optimise compression to the given set of circumstances. The adaption method may, depending on the factors, adjust the luminance of the content as to provide the highest amount of visible range at the display, or optimise the compression to the scene-referred representation of the data. This metadata then may include a normalisation factor, or parameters to reshape the transfer function for optimal compression. The framework provides the facility to evaluate the decompressed output for quality either in terms of base-band coding errors only, or with video encoding included.

5.4 Optimisation Method

For an adaptive HDR video compression method to be effectively implemented, a pipeline must be designed to provide the necessary structure for data processing.

The pipeline provides the necessary stages to allow the uncompressed HDR video frames to be passed in, for the compression method to process the frames and adapt accordingly, for the video bit-stream and required metadata be packaged in a container for distribution, for the bit-stream and metadata to be decoded to produce the decompressed HDR video frames, and optimally for the quality of the compression method to be tested.

Figure 5.5 presents the pipeline for PTF_a. The optimisation is seeded with an initial γ parameter for testing, this is compressed with PTF and then optionally encoded and decoded using a video encoder. The use of a video encoder, such as x265, HM [SOH+12],

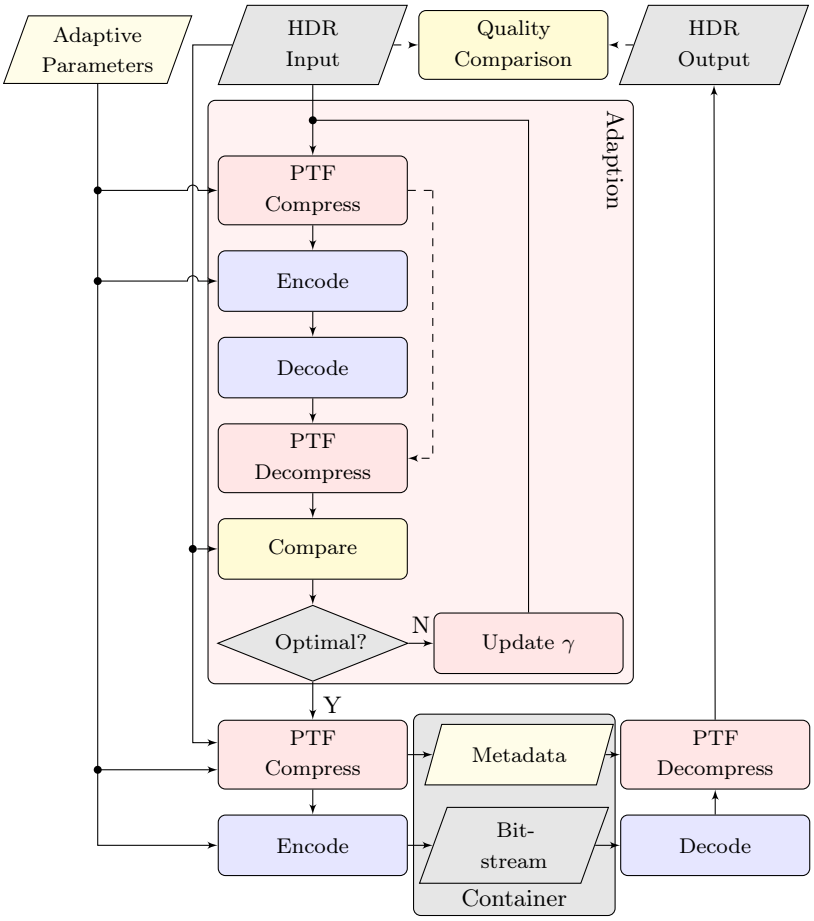


Figure 5.5: Adaptive compression pipeline for PTF_a allowing iterative optimisation of PTF to achieve the highest quality compression. Dashed lines denote optional steps, such as the coding only optimisation that avoids the encode and decode optimisation steps.

or VP9 [MBG+13], and decoder is optional as it incurs a significant performance penalty on optimisation. If a video encoder is not included however then the selected γ may be sub-optimal. This is due to the video encoder and associated parameters such as QP having an effect on the ideal γ value, as shown in Section 5.2. The data is then decompressed with PTF and compared with the input to decide whether the chosen γ value is optimal. If the value is not optimal then a new γ value is chosen and then next iteration is performed. When the optimal value is found, it is used with PTF to compress the data before encoding it with the video encoder and distributed as a single video bitstream and accompanying metadata in a container format. This package can then be passed to a video decoder and PTF decompression function for display. If additional quality metric are desired then they can be calculated using the decoded HDR output frames. The metadata required for PTF_a includes the normalisation factor $\mathfrak{N}$ as with PTF_γ however it now must include the adaptive γ value chosen.

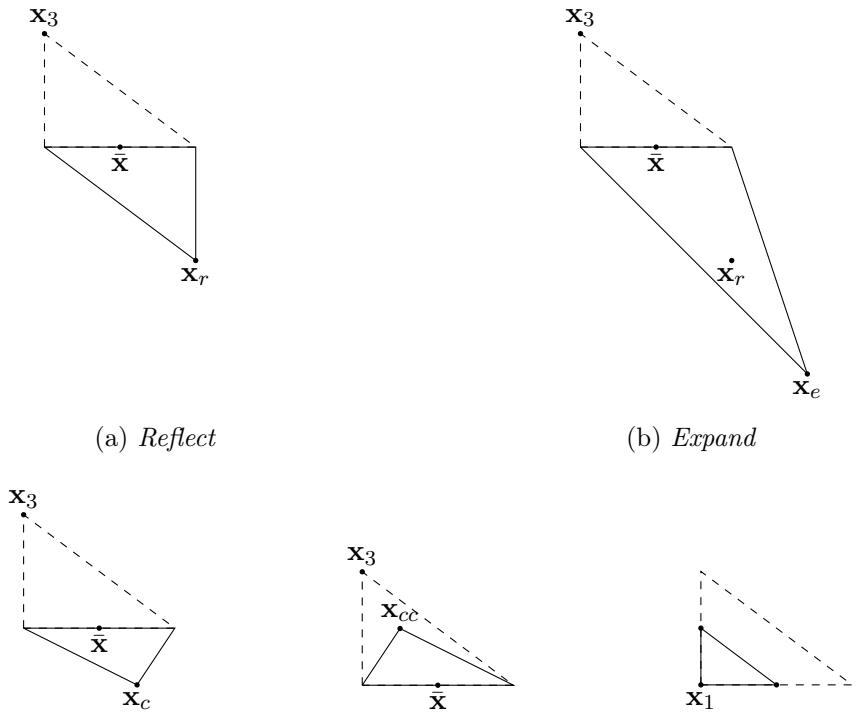
The optimisation procedure can be performed per frame however, adjusting the parameter too frequently may reduce compression efficiency as it results in global frame adjustment. Alternatively it can be adjusted at the intra-frame period, per scene or per sequence.

The choice of algorithm to select the optimal γ parameter is not restricted, however it is crucial to the performance of PTF_a. In this implementation and the evaluation in Section 5.5, the Nelder-Mead downhill simplex method for function minimisation [NM65] was used as implemented by MATLAB in the `fminsearch` function. It was chosen due to it's lack of a requirement for explicit function derivatives [LRW+98] , unlike other methods which require an analytical expression of the function gradient which cannot be derived for image quality. The Broyden-Fletcher-Goldfarb-Shanno (BFGS) Quasi-Newton non-linear function minimisation method [Bro70; Fle70; Gol70; Sha70] was also tested, however it converged on less optimal γ parameters than Nelder-Mead.

5.4.1 Nelder-Mead Function Minimisation

The Nelder-Mead method is an algorithm for minimising a real-valued objective function $f(x)$ for $x \in \mathbb{R}^n$ where n is the dimensionality of the search space. The method creates a simplex in n dimensions which it then steps one vertex at a time towards a minima in the objective function. Four scalars must be specified corresponding to *reflection* (ρ), *expansion* (χ), *contraction* (γ), and *shrinkage* (σ). The parameters should satisfy:

$$\rho > 0, \quad \chi > 1, \quad \chi > \rho, \quad 0 < \gamma < 1, \quad 0 < \sigma < 1. \tag{5.2}$$



(a) *Reflect* (b) *Expand*

(c) *Outside contract* (d) *Inside contract* (e) *Shrink*

Figure 5.6: Nelder-Mead simplex transformations in 2d after different stages of the algorithm have been performed. Dashed line denotes the kth iteration simplex while the solid line denotes the simplex for the iteration, $k+1$.

The values for the parameters in the MATLAB implementation of Nelder-Mead are

$$\rho = 1, \quad \chi = 2, \quad \gamma = \frac{1}{2}, \quad \sigma = \frac{1}{2}, \tag{5.3}$$

these values are taken to define the *standard* Nelder-Mead algorithm.

At the kth iteration the algorithm defines a simplex Δ_k with it's $n+1$ vertices, $\mathbf{x}_1^{(k)}, \ldots, \mathbf{x}_{n+1}^{(k)}$, each of which exist in $\mathbb{R}^n$. The vertices are ordered such that

$$f_1^{(k)} \leq f_2^{(k)} \leq \cdots \leq f_1^{(k)}, \tag{5.4}$$

where $f_1^{(k)}$ denotes $f(\mathbf{x}_i^{(k)})$. The kth iteration produces a set of $n+1$ vertices to define a new simplex for the $k+1$th iteration. As the aim is the minimise f, $\mathbf{x}_{n+1}^{(k)}$ is the *worst* vertex, $\mathbf{x}_n^{(k)}$ is the *next-worst* vertex, while $\mathbf{x}_1^{(k)}$ is the *best* vertex. Each iteration k of the algorithm produces either a new vertex which replaces $\mathbf{x}_{n+1}^{(k)}$ in the set or vertices for the iteration $k+1$, or a new set of n points which, together with $\mathbf{x}_1^{(k)}$ form the complete simplex for the iteration $k+1$.

The stages of a single iteration k of the algorithm are presented as follows

1. **Order:** Sort the vertices to satisfy $f_1^{(k)} \leq f_2^{(k)} \leq \cdots \leq f_1^{(k)}$, using tie breaking rules if required.

2. **Reflect:** Compute the *reflection point* $\mathbf{x}_r^{(k)}$ from

$$\mathbf{x}_r^{(k)} = \bar{\mathbf{x}}^{(k)} + \rho\left(\bar{\mathbf{x}}^{(k)} - \mathbf{x}_{n+1}^{(k)}\right), \tag{5.5}$$

   where $\bar{\mathbf{x}}^{(k)} = \sum_{i=1}^{n} \mathbf{x}_i^{(k)}/n$ is the centroid of the n best vertices. If $f_1^{(k)} \leq f_r^{(k)} < f_n^{(k)}$, return the *reflection point* $\mathbf{x}_r^{(k)}$ and terminate the iteration.

3. **Expand:** If $f_r^{(k)} < f_1^{(k)}$, calculate the *expansion point* $\mathbf{x}_e^{(k)}$,

$$\mathbf{x}_e^{(k)} = \bar{\mathbf{x}}^{(k)} + \chi\left(\mathbf{x}_r^{(k)} - \bar{\mathbf{x}}^{(k)}\right) = \bar{\mathbf{x}}^{(k)} + \rho\chi\left(\bar{\mathbf{x}}^{(k)} - \mathbf{x}_{n+1}^{(k)}\right). \tag{5.6}$$

   If $f_e^{(k)} < f_r^{(k)}$, return the *expansion point* $\mathbf{x}_e^{(k)}$ and terminate the iteration; otherwise $f_e^{(k)} \geq f_r^{(k)}$, accept the *reflection point* $\mathbf{x}_r^{(k)}$.

4. **Contract:** If $f_r^{(k)} \geq f_n^{(k)}$, perform a *contraction* between $\bar{\mathbf{x}}^{(k)}$ and the smaller of $\mathbf{x}_{n+1}^{(k)}$ and $\mathbf{x}_r^{(k)}$.

   (a) **Outside:** If $f_n^{(k)} \leq f_r^{(k)} < f_{n+1}^{(k)}$ perform an *outside contraction*

$$\mathbf{x}_c^{(k)} = \bar{\mathbf{x}}^{(k)} + \gamma\left(\mathbf{x}_r^{(k)} - \bar{\mathbf{x}}^{(k)}\right) = \bar{\mathbf{x}}^{(k)} + \gamma\rho\left(\bar{\mathbf{x}}^{(k)} - \mathbf{x}_{n+1}^{(k)}\right). \tag{5.7}$$

   If $f_c^{(k)} < f_r^{(k)}$, return $\mathbf{x}_c^{(k)}$; otherwise perform a shrink.

   (b) **Inside:** If $f_r^{(k)} \geq f_{n+1}^{(k)}$, perform an *inside contraction*

$$\mathbf{x}_{cc}^{(k)} = \bar{\mathbf{x}}^{(k)} - \gamma\left(\bar{\mathbf{x}}^{(k)} - \mathbf{x}_{n+1}^{(k)}\right). \tag{5.8}$$

   If $f_{cc}^{(k)} < f_{n+1}^{(k)}$, return $f_{cc}^{(k)}$; otherwise perform a shrink.

5. **Shrink:** Calculate a new set of n vertices $\mathbf{v}_i^{(k)} = \mathbf{x}_1^{(k)} + \sigma\left(\mathbf{x}_i^{(k)} - \mathbf{x}_1^{(k)}\right)$, $i = 2, \ldots, n+1$. The vertices of simplex Δ_{k+1} consist of $\mathbf{x}_1^{(k)}, \mathbf{v}_2^{(k)}, \ldots, \mathbf{v}_{n+1}^{(k)}$.

The implementation in MATLAB of Nelder-Mead is as described in [LRW+98] and uses the following tie-breaking rules when vertices result in equal function values. The rules assign the new vertex the highest possible position in the evaluation ordering.

Non-shrink ordering rule. During a non-shrink step, the vertex $\mathbf{x}_{n+1}^{(k)}$ is discarded. The new vertex, $\mathbf{v}^{(k)}$, assumes the position $j+1$ in the vertices of Δ_{k+1}, where

$$j = \max_{0 \leq \ell \leq n} \left\{ \ell \mid f(\mathbf{v}^{(k)}) < f(\mathbf{x}_{\ell+1}^{(k)}) \right\}; \tag{5.9}$$

the remaining vertices retain their previous ordering.

Shrink ordering rule. During a shrink step, the only vertex preserved from Δ_k is $\mathbf{x}_1^{(k)}$. If one or more of the new are tied as the best point then retain $\mathbf{x}_1^{(k)}$ as the best point: if

$$\min \left\{ f(\mathbf{v}_2^{(k)}), \cdots, f(\mathbf{v}_{n+1}^{(k)}) \right\} = f(\mathbf{x}_1^{(k)}), \tag{5.10}$$

then $f(\mathbf{x}_1^{(k+1)}) = f(\mathbf{x}_1^{(k)})$. Figure 5.6 shows the simplex transformations performed by Nelder-Mead at each stage of the algorithm.

For practical use Nelder-Mead requires an upper bound on the number of iterations performed to ensure the algorithm terminates in a finite period. Three different tests are performed, $term_x, term_f, fail$; in normal usage the resulting disjunction taken to be whether to terminate the algorithm.

1. $term_x$: the *domain* convergence test evaluates to true when the volume of the simplex Δ_k reduces past a certain threshold, tol_x.

2. $term_f$: the *function* convergence test evaluates to true when the values from all the function evaluations converge to within a certain threshold, tol_f.

3. $fail$: the *no-convergence* test becomes true when the number of iterations or the number of function evaluations exceeds a certain threshold, num_k and num_f.

Common implementations define $term = term_x \vee term_f \vee fail$ however MATLAB uses $term = (term_x \wedge term_f) \vee fail$. The thresholds used for this evaluation were the MATLAB defaults, $tol_x = 1 \times 10^{-4}, tol_f = 1 \times 10^{-4}, num_k = 200 * n, num_f = 200 * n$.

Nelder-Mead was chosen for this implementation of PTF_a partially because, of the two optimisation methods tested, it resulted in the highest objective quality but also because it is a comparatively fast optimisation method. Other than during a *shrink* step the algorithm only needs one or two function evaluations per iteration. As the function to optimise is expensive to compute this is a significant benefit. Furthermore is has been shown experimentally that shrink steps almost never occur in practice. Torczon reports only 33 shrink steps in 2.9 million iterations on a set of general test functions [Tor89]. The rarity of shrink steps is also confirmed by Lagarias et al. [LRW+98]. The final justification for the choice of optimisation function is that of the function minimisation

algorithms available in MATLAB, Nelder-Mead optimised the γ parameter to result in the highest objective quality. Other optimisation functions may result in more optimal values for γ and this could be the study of future work.

Following the use of this algorithm in 1 dimension the value a in PTF_a is set to a local minima of the search domain, $\min_a f(\text{PTF}'_a(c(\text{PTF}_a(x))), x)$, where x is the video sequence to be compressed, c is optionally a function to encode and decode a video bit-stream, and $f(x, y)$ is the image comparison function. In this evaluation $f(x, y) = \frac{1}{\text{HDR-VDP}(x,y)}$ or $f(x, y) = \frac{1}{\text{puPSNR}(x,y)}$, depending on which metric was being optimised for. However, in general, f can be any image quality comparison function which can be minimised to yield improved quality.

5.5 Results

For adaption to be considered useful it must result in a higher objective quality than that available from a fixed TF compression method. The primary focus of these results is to evaluate the quality as a result of reshaping the transfer function. Therefore, the scene-referred luminance and a number of display-referred luminances were selected for testing to simulate the luminances required for a range of ambient environments, this is the independent variable, *luminance*.

This evaluation was performed on a dataset of 147 images. As the focus is on adaption to content, a dataset of images was chosen to gain an understanding of how adaption performs under a wide range of content. The dataset initially contained scene-referred data however it was also compressed to a range of luminances using a 99th percentile grading method. The values for the independent variable, *luminance*, chosen were 1,000, 2,000, 4,000 and 10,000 cd/m^2 as they represent a variety of typical values for display-referred data, and the scene-referred representation. The dataset contained images from a variety of sources with a range of content and capture methods. 105 of the 106 images from the Fairchild dataset [Fai07] were used, calibrated where possible using the provided data and rescaled to 1920×1080 using a bilinear resampling method. 19 of the 20 images presented in Appendix A were also tested[2]. 12 frames from the sequences provided to MPEG for the HDR and Wide Colour Gamut Call for Evidence [LFH15] were tested. Frames from Blender's Tears of Steel [Hub12]; Stuttgart Media University's EXR Sequences [FGE+14] Showgirl 1 & 2, Cars Fullshot, Fireplace 2; Technicolor's Seine [LLF13]; Warwick University's Jaguar, and CGRoom were used. The mean peak value for the scene-referred data was 66,059 cd/m^2, the mean value was 1,347 cd/m^2, and the mean dynamic range was 25.10 stops.

[2]Memorial was excluded on account of its resolution.

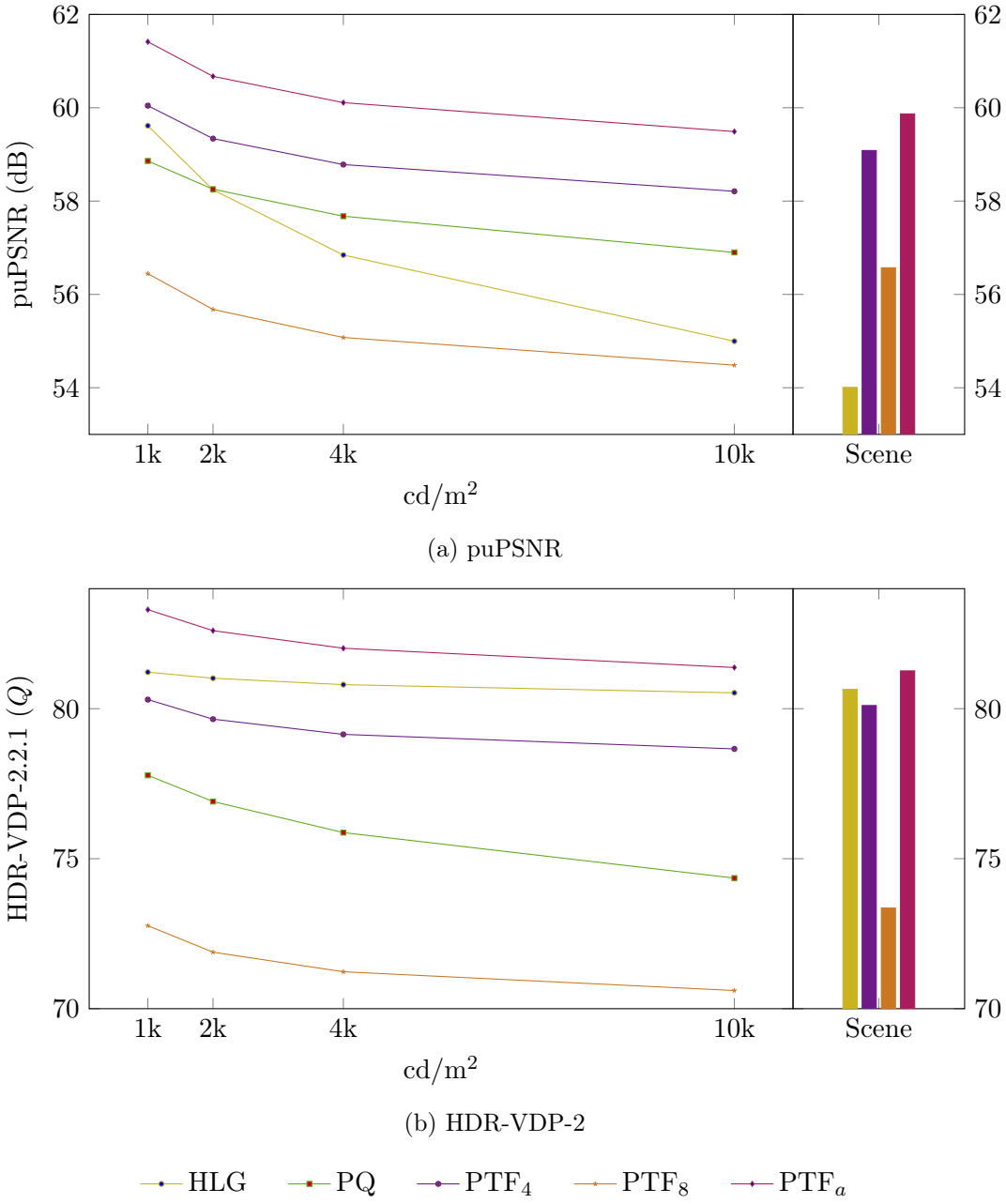
The $Y'C_BC_R$ colourspace was used with the correct primaries for the content and 4:2:0 chroma-subsampling. Video encoding was not used for this evaluation primarily to gain an understanding of adaption at the base-band level. Secondly because encoding images with a video encoder is not necessarily indicative of how the encoder will perform on a video sequence as the image will be encoded as an I-frame. Thirdly, to reduce the computational cost of the evaluation as many iterations of the encoding function are needed to optimise the compression method.

5.5.1 Quality

Figure 5.7 presents the average results of PTF_a after optimisation when compared with HLG, PQ, PTF_4, and PTF_8 for both HDR-VDP-2 and puPSNR. For both metrics, a higher result indicates less distortion in the image, or fewer artefacts. These two metrics were chosen as they correlate most highly with a subjective evaluation [MDB+16a]. In the evaluation the optimisation comparison function f was chosen to be the reciprocal of $\text{puPSNR}(ref, dec)$ where ref is the set of reference frames, dec is the set of decoded frames from $\text{PTF}'_a(\text{PTF}_a(ref))$, $\text{PTF}'_a(x)$ is the PTF decompression function. The Nelder-Mead optimisation method was seeded with an initial value of 4 as PTF_4 was found to give the highest objective results in Chapter 3. Seeding the algorithm with random starting locations between 1 and 12 was tested, however it resulted in a lower quality result. This is likely due to the optimisation algorithm finding local minima rather than the global minima. Figures 5.1 and 5.2 however appear to show no local minima, it is hypothesised that the granularity of the brute force evaluation was not sufficient to resolve the local minima present.

Figure 5.7a shows the results for puPSNR. On the display-referred data, the ranking of HLG dropped from 3rd to 4th as *luminance* increased to 10,000 cd/m^2, the other methods order remained constant with regards to each other and was: PTF_a, PTF_4, PQ, PTF_8; and the following order for scene-referred data: PTF_a, PTF_4, PTF_8, HLG, PQ. Except for HLG, the difference in quality for display-referred data remained constant between the methods but dropped as *luminance* increased. The methods saw an increase in objective quality when tested using scene referred data over data graded to 10,000 cd/m^2.

Figure 5.7b shows the objective results for the HDR-VDP-2 metric, the compression methods are ranked in the following order from highest quality to lowest quality for the display-referred data: PTF_a, HLG, PTF_4, PQ, PTF_8; and the following order for the scene-referred data: PTF_a, HLG, PTF_4, PTF_8, PQ. The difference in quality drop remained fairly constant as the *luminance* of the display-referred data was increased, except for HLG where quality was almost preserved. On scene-referred data the top



(a) puPSNR

(b) HDR-VDP-2

Figure 5.7: Coding performance in (a) puPSNR and (b) HDR-VDP-2 for PQ, HLG, PTF_4, PTF_8 and PTF_a when tested on a dataset of 147 images graded at a range of display-referred luminances and the scene-referred luminance. The results for PQ on scene-referred data are not reported as the method is limited to representing data of 10,000 cd/m^2 or less.

three methods, PTF_a, HLG, and PTF_4 produced similar results whereas PTF_8 and PQ produced lower results. PTF_a and PQ performed worse than the 10,000 cd/m^2 result however the other methods performed better.

5.5.2 Performance

Luminance (cd/m^2)	**Optimal γ**	**Optimisation Time (s)**
1,000	3.25	108.75
2,000	3.34	109.04
4,000	3.40	107.28
10,000	3.48	107.30
Scene	4.22	107.67

Table 5.2: Mean trained γ values along with optimisation time in seconds for PTF_a on the 147 image dataset used in the evaluation. The mean luminance of the scene-referred dataset is 66,059 cd/m^2.

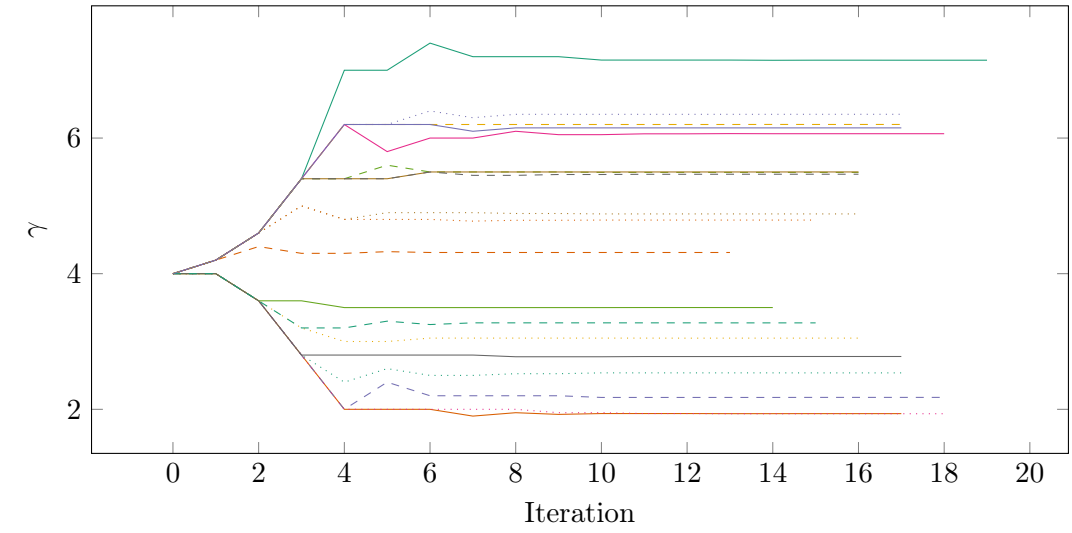
Table 5.2 shows the resulting mean adaptive γ values along with the mean optimisation time from training PTF_a on the dataset used for the evaluation for each *luminance*. The mean trained γ value increased slightly as the display-referred *luminance* increased but jumped up significantly for scene-referred data. The mean optimisation time was roughly constant at around 1m 49s per image.

5.5.3 Optimisation

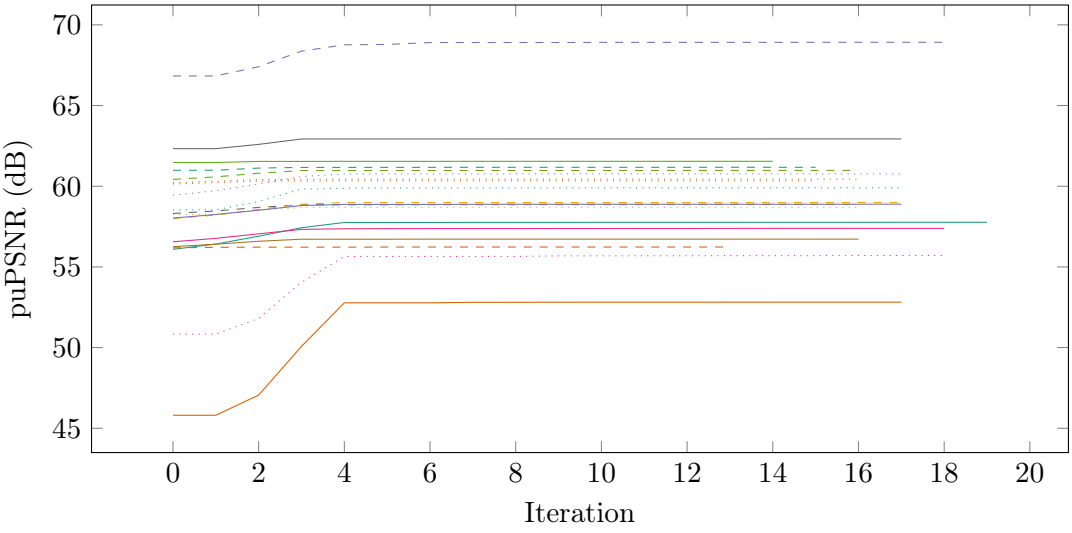
Figure 5.8 shows how γ and puPSNR change during the course of the optimisation. Figure 5.8a shows how the optimisation is initialised at a γ value of 4 and then quickly diverges as the search algorithm finds higher performing γ values for individual images on each iteration. Figure 5.8b shows the gain in puPSNR improves as iteration. Each image starts with a baseline quality, the quality of the image when encoded with PTF_4, which is then improved as the search algorithm finds higher performing γ values. Particularly striking is the quality improvement achieved on B (10) by the optimisation. Initially, the puPSNR is 45.81 with a γ value of 4, however this climbs to 52.82 puPSNR when its optimal γ of 1.94 is achieved.

5.6 Analysis & Discussion

PTF_a was trained using the puPSNR metric so it is expected that the method performs at least as well as PTF_4, the second highest performing method. The result of the



(a) γ



(b) puPSNR

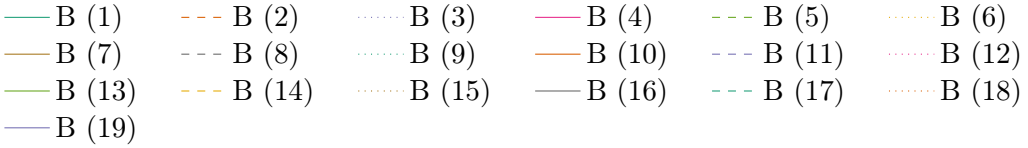


Figure 5.8: Example optimisations of PTF_a for 19 images from the Warwick dataset shown in Appendix A when graded to scene referred *luminance*. (a) shows how γ changes during each iteration of an image. (b) shows the gain in puPSNR during each iteration of an image.

optimisation was that PTF_a produced the highest objective result at all values of *luminance* including scene-referred data. PTF_a exhibited a fairly consistent gain of 1.33 puPSNR (2.20%) over the next best method, PTF_4, and 2.52 puPSNR (4.14%) over PQ on the display-referred data. PTF_a exhibited a gain of 0.80 puPSNR (1.30%) over PTF_4 on scene-referred data.

It should be noted however that the training did not overfit PTF_a to the puPSNR metric. PTF_a outperformed the other tested methods in terms of HDR-VDP-2 at all values of *luminance*, gaining on average 1.43 VDP (1.74%) over HLG, the second highest performing method, for display-referred data and 0.62 VDP (0.76%) for scene-referred data. PTF_a gained 2.88 VDP (3.50%) over PTF_4 on display-referred data and 1.16 VDP (1.42%) on scene-referred data. This result shows that PTF_a optimisation generalises and adds robustness to the method. The results show that by adapting the γ parameter in PTF_a, an increase in both HDR-VDP-2 and puPSNR can be achieved over the use of PTF with a fixed γ parameter.

It should be noted that the performance of PQ was severely hampered on scene-referred data. This is because PQ uses a fixed normalisation value of 10,000 cd/m^2 and is unable to represent data above this limit without the perceptual curve being reformulated.

From Table 5.2 it can be seen that as the *luminance* increases the mean γ value increases. This correlates with what is known about the HVS and visual perception. As the luminance of the content increases through the mesopic range to the photopic range the perceptual response moves from a ½ power function as described by the de Vries-Rose relationship [Kel77] to the logarithmic response described by the Weber-Fechner relationship [SYD87]. The *luminance* values were chosen to reflect typical values for the peak luminance of displays. This suggests a relationship between the compression curve and the peak luminance of the display which can be exploited by PTF_a to improve compression efficiency.

Table 5.2 presents the mean time taken to optimise PTF_a at each luminance. This evaluation was performed on single frames and without the use of a video encoder although a high performance version was not implemented and instead the optimisation was performed in MATLAB. A mean time for all values of *luminance* was 108 s however, this limits the practicality of direct optimisation for PTF_a. If the optimisation was performed on video sequences and using high quality video encoding the time taken to optimise would be greatly increased. The use of optimisation also prevents the method from being used in a real-time environment as such a system is required to process frames at the capture frame-rate, usually 30 or 60 fps.

Care also needs to be taken when optimising for video sequences as large jumps in

encoded γ will reduce compression efficiency and could introduce temporal artefacts. A potential solution to this problem would be to limit the range of each successive frame's γ to be within a certain distance of the previous frame's γ. This also would have the benefit of reducing the search space and therefore reducing the time needed to optimise PTF_a.

The benefit of adapting a single γ parameter over the use of a LUT as used by other methods is the reduced metadata requirement. Lu et al. [LPY+16b] proposed approximating the LUT with a piece-wise polynomial curve to reduce the size of the reshaping metadata however they then had to use the inverse of the polynomial to apply the forward reshaping to the data, increasing complication.

5.7 Conclusion & Future Work

This chapter has presented PTF_a, an adaptive compression method based on power functions and an extension to PTF.

It provided the motivation for adaptive compression, taking the conclusion to Chapter 4 to show why content may need to be adjusted for the ambient environment. It demonstrated, using a brute-force evaluation, how content compresses under a range of γ values and showed that varying the γ based on the content could lead to significant improvements in compression quality.

A generic framework for adaptive compression was introduced taking into account parameters from the display, such as the peak luminance, the environment at the display, most notably the ambient illumination, and the content. The framework allowed these parameters to feed into the HDR compression method with the goal of providing an improved experience. This chapter then explained in detail the method used in PTF_a for optimising the γ parameter based on the content, including a thorough explanation of the Nelder-Mead method used in the following evaluation for function optimisation. The flow and operation of PTF_a was presented in a processing pipeline which also served as the experimental framework.

This chapter then presented the results of an experiment to determine whether adapting the compression method to the content has a positive impact on the quality. This was followed by a thorough analysis of the results and their impact.

The results of the evaluation show that PTF_a produces higher quality compression than PTF_4, PTF_8, PQ [MND13] and HLG [Bor14], when tested with both puPSNR and HDR-VDP-2. It should be noted that adaptive compression enhanced quality under a different metric than the one used for the optimisation process. This provides important evidence towards the robustness of adaptive compression and demonstrates that the

optimisation did not simply tune the γ parameter to the metric.

The results also show that the ideal γ value for the compression of the content is dependent on the luminance of the display intended for the presentation of that content. Chapter 4 has shown that the peak luminance required to perceive the detail in the content is dependent on ambient environment. We therefore have a system where adjusting the luminance of the content such that the range is visible on the display requires that the γ value is in turn adjusted to ensure optimal compression.

The research question posed in the introduction to the chapter is then answered. This chapter has demonstrated that compression performance can be improved by adapting the compression method to the content, the ambient environment and the luminance of the display.

The computation required to optimised even a single parameter incurs a significant penalty on the performance of the method, even with the high computational efficiency of PTF. It is possible that a closed form solution to the γ selection is possible. The proposed method could be implemented for HDR video on either a per-frame or per-sequence basis, with per-frame having far greater performance, however the computational expense incurred renders this method incapable of achieving real-time performance. Chapter 6 will look into the potential to develop a model for predicting optimal γ values with greatly reduced computational demand.

A topic for future study will investigate the effects of the video encoder and its parameters on adaption. Section 5.2 has provided evidence that the QP has an affect on the optimal γ value. However, the computation required to optimise high-quality encoded video sequences restricts what can be achieved in a reasonable time-scale.

Chapter 6

Deep Learning for Adaptive Compression

Chapter 5 introduced PTF_a as a method for adaptive HDR compression, and showed that adapting the γ value to the content can yield improvements in objective quality over a fixed transfer function such as PTF_4, PQ or HLG. The method of γ value selection used in PTF_a was an iterative search for the optimal γ value per-sequence or per-frame using a function minimisation algorithm. This iterative process requires many evaluations of an objective function which, in the case of the compression of a video sequence, is very computationally expensive, taking, on average, 1m 49s per frame with an unrestricted search.

This chapter will introduce two methods to improve the computational performance of Adaptive PTF compression. Firstly, this chapter will present a regression method based on simple image statistics to predict γ values. A regression model is computationally and memory efficient, however it is restricted to relying on the features provided as input. This chapter will show that a regression model can predict γ values with relatively high accuracy. However, the results show that the predictions yield compression with a lower objective quality than that achieved by using a constant γ value in the form of PTF_4 and the method is therefore not worthwhile.

Secondly, as a regression model cannot improve the quality of PTF_4, this chapter presents a deep learning method using Convolutional Neural Networks (CNNs) to train a model. The CNN is able to learn image features and use them to predict the γ values needed for high quality compression. An evaluation demonstrates that the CNN method produces higher quality compression than PTF_4, and almost matches the quality at PTF_a but with orders of magnitude less computation. The aim of this chapter is to accelerate optimised compression to a rate that enables real-time compression while

using an adaptable γ parameter to achieve high-quality compression.

6.1 Regression Model

Fitting a model to image statistics, the independent variables, using regression techniques allows for a computationally and memory efficient forward-pass when predicting the dependent variable, γ in this system. Image statistics can be collected in a computationally efficient manner requiring few passes over the data. The statistics collected were: min, max, mean, median, harmonic mean, standard deviation, kurtosis, $\log_{10}$ mean, exp mean log, weber contrast, dynamic range, image key, skew, interquartile range, geometric mean and log kurtosis. The resulting model is memory efficient and can be represented using a small number of parameters.

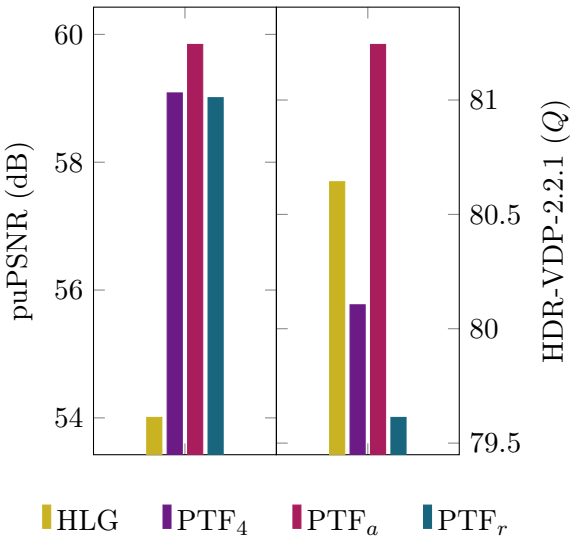


Figure 6.1: Scene-referred, coding-only puPSNR and HDR-VDP-2 results for the regression model on the 147 images dataset. The regression model was used to predict γ values for PTF with the results compared against HLG, PTF_4 and PTF_a.

The dataset used to perform the regression was the 147 image dataset at scene-referred luminance only. The regression is an offline process to build the model and takes the dataset of HDR images and pairs of γ values as input. k-fold cross-validation was used with 146 images used to train and one image used to validate. The resulting model used the Kurtosis image statistic only and produced an mean absolute error of 1.06.

6.1.1 Results

Figure 6.1 shows the results for the regression model, PTF_r. The model was compared with HLG, PTF_4 and PTF_a on the scene-referred data using a coding-only compression scheme to 10-bit $\text{Y}'\text{C}_\text{B}\text{C}_\text{R}$ 4:2:0. The results show that PTF_r does not perform as well as PTF_a or PTF_4 when tested with either puPSNR or HDR-VDP-2. PTF_r does outperform HLG when tested with the puPSNR metric however it does not when tested with the HDR-VDP-2 metric. PQ results are not shown as the method is not suitable

for use with scene-referred data.

6.1.2 Analysis & Discussion

The reason for the poor performance of the regression model based implementation of PTF_r is that the predicted γ values are too inaccurate to result in high-quality compression. This is shown by the error resulting from the training being 1.06. The predicted γ values must more closely match the ideal values from the optimisation process for regression to be considered a useful adaptive method. For γ prediction to be a useful technique, the resulting quality must be higher than when using a constant γ value, such as in PTF_4. This chapter will now show that a CNN is able to predict γ values from content with higher accuracy as the necessary features are not predetermined, but can instead be extracted during the training process.

6.2 Deep Learning Model

This section will investigate whether training a CNN to learn relevant features from the image will result in predictions of the dependent variable, γ, with lower error. The independent variable used is the HDR image itself. A CNN was chosen as it is capable of learning the features it needs to reduce the error, this is achieved by adjusting the convolution performed at each layer.

6.2.1 Method

A (Deep) Neural Network consists of layers of non-linear transformations. The input, x_l, to each layer, l, is linearly transformed by a matrix of learnable weights, W_l, and biases, b_l, before each dimension is non-linearly transformed by an activation function, $f(a)$. The output from each layer, $o_l = f(W_l x_l + b_l)$, is used as the input to the next layer, $x_{l+1} = o_l$, in a feed-forward manner. The first layer, where data is presented to the network, is called the input layer, and the final layer, where data is extracted from the network, is called the output layer. All layers between the input and output layers are called hidden layers.

CNNs contain specialised layers for image processing which utilise convolutional kernels to extract spatial correlations in the image. This kernel can again be expressed as a linear transformation matrix, W_l, but can be highly sparse and contain repeated weights. The activation function chosen for the network in this chapter is the non-saturating Rectified Linear Unit (ReLU) [HSM+00], $f(a) = \max(0, a)$, which is commonly used in CNNs [LBH15] as it is computationally efficient and can allow for faster and more effective training on complex datasets [GBB11]. CNNs frequently contain subsampling

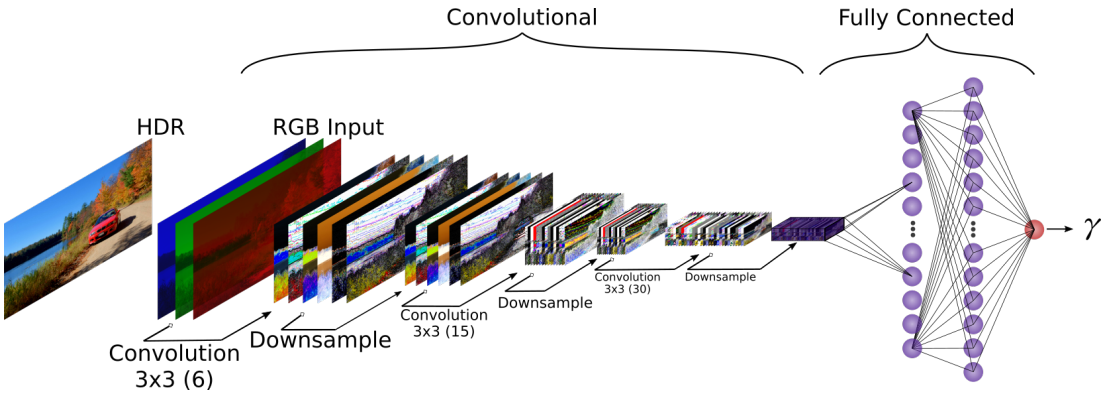


Figure 6.2: Structure of the Convolutional Neural Network used in this chapter showing the convolutional and fully-connected stages. A HDR image is the sole input to the network and predicted γ value the sole output.

layers to lower the dimensionality of the data and so to reduce computational cost. Max pooling is one method of subsampling which returns the maximum of non-overlapping $N \times N$ regions of the input and reduces each dimension by a factor of N.

The weights and biases of each layer are learned from the data using a supervised technique such that the chosen loss function is minimised. In this method the loss function was the L1-loss between the network output and supervisory signal. The loss is minimised using stochastic gradient descent (SGD) and the backpropagation algorithm. The algorithm uses the chain rule to compute the gradient of the loss with respect to the network parameters. The parameters are then updated in the direction of the gradient so as to minimise the loss.

The input to the CNN in this chapter is a full resolution HDR image. It is passed through three sets of convolutional layers, each followed by max-pooling layers. Each convolution consists of a 3×3 kernel with a stride of 2. The first layer is comprised of 6 filters, followed by 15 filters for the second layer and 30 for the third layer. Each max pooling layer uses a 2×2 region, halving the resolution of the data in each dimension. The output from the three convolutional layers is a high-level description of the image content. The output of the convolutional layers is used as the input to the fully-connected network with an input layer of size 64 and on to a hidden layer of size 512. The output layer produces a single number, the predicted γ value used for compression. The architecture of the network is presented in Figure 6.2.

The dataset used to develop the network is the same as used in Chapter 5. It consisted of 147 images graded at five luminance levels, four display-referred luminance levels (1,000, 2,000, 4,000 and 10,000 $\mathrm{cd/m^2}$) and the scene-referred luminance, which were paired with their optimal γ values computed in Chapter 5 to give a total of 735

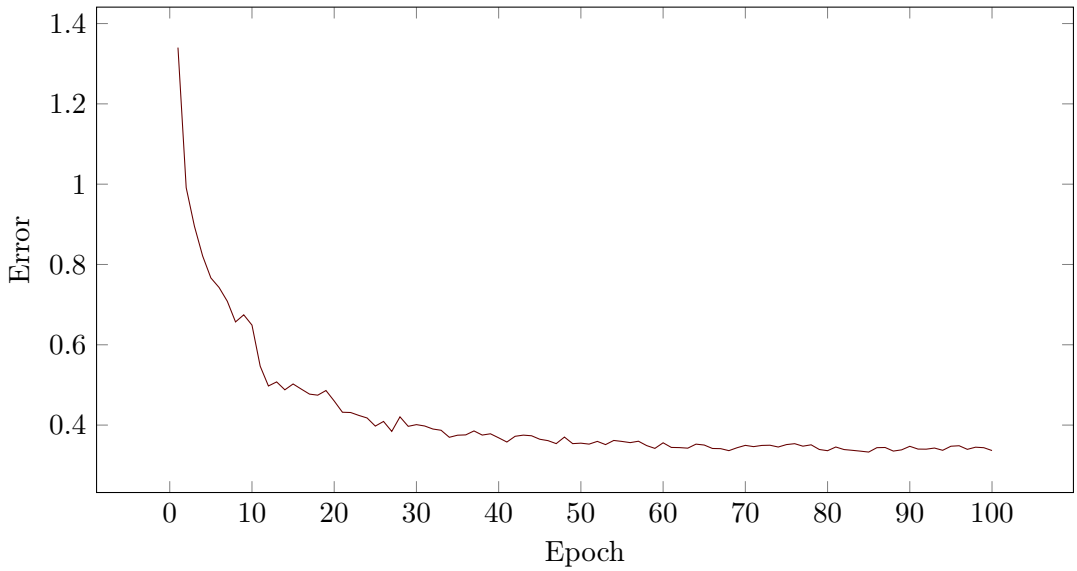


Figure 6.3: PTF_p training convergence over the 100 epoch training time. The error can be seen to drop every 10 epochs as the learning rate is adapted.

pairs. From this dataset, 661 pairs were used as the training set and 74 pairs were reserved for testing the network. The network was trained using PyTorch for 100 epochs using the ADAM optimiser [KB14] with the default β_1 and β_2 parameters. The learning rate for SGD halved every 10 epochs from an initial value of 0.0001. The network took 9 hours, 57 minutes to train using an Nvidia GeForce GTX1080. After training the network, the mean absolute error on the test set was 0.303. Figure 6.3 shows the convergence of the network during training.

Figure 6.4 shows the general data flow for PTF_p. The training is an offline process using backpropagation to set the best possible weights for the network, and like regression takes the dataset of HDR images and γ pairs as an input. The many weights and biases are passed to the production CNN for feed-forward operation. The HDR input frame is then passed to the CNN for the convolutional stage and then the many learned features are passed to the fully-connected stage to estimate the optimal γ value. This value is then used by PTF to compress the HDR input frame. $\text{Y}'\text{C}_\text{B}\text{C}_\text{R}$ data with possible chroma sub-sampling is passed to the video encoder to create a bit-stream for distribution and the adaptive γ value used for compression is packaged alongside the bit-steam in a container format. Decoding follows a similar process to PTF_a where the γ parameter is taken from the metadata and used to decompress the decoded bit-stream from the video container. Finally, the quality between the reference HDR input and the distorted HDR output frames can be calculated by a metric such as puPSNR or HDR-VDP-2 to give an overall quality score.

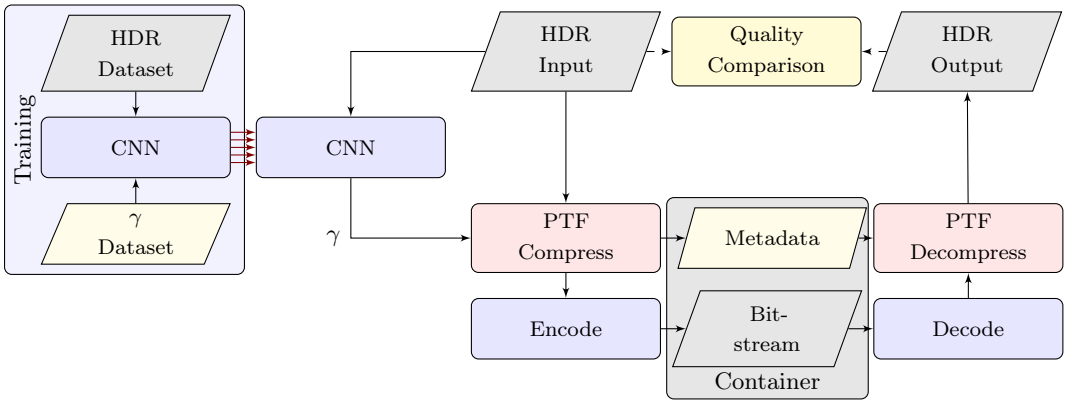


Figure 6.4: Pipeline showing the operation and dataflow used in PTF_p. A set of red arrows ⇉ denotes the transfer of weights and biases from the training network to the feed-forward network.

6.2.2 Results

The mean absolute error between optimal γ values, calculated in Chapter 5, and the predicted γ values was 0.303. To understand how this affects perceptual results it was necessary to use the predicted γ values to compress data. PTF_p was evaluated against 3 state-of-the-art one-stream methods, PQ [MND13], HLG [Bor14] and PTF_4, and also against PTF_a, which was presented in Chapter 5, to see the loss caused by employing a CNN. The methods were tested using data graded at the display-referred luminances of 1,000, 2,000, 4,000 and 10,000 cd/m^2 and the scene-referred luminance. The content was stored in PFM frames as single-precision floating-point data were needed to store the high values needed for scene-referred data. The pipeline was executed in its entirety on each image and at the scene and display-referred luminances tested. The dataset contained 74 images randomly selected from the 735 image dataset consisting of 147 images graded at 5 luminances. The testing set contained 21 images graded at 1,000 cd/m^2, 11 at 2,000 cd/m^2, 19 at 4,000 cd/m^2, 7 at 10,000 cd/m^2 and 16 at scene-referred luminance, these numbers arose as a random permutation of the dataset which was concatenated to augment the data for training.

The methods were compared for perceptual quality on an objective basis using the puPSNR and HDR-VDP-2 metrics. The performance of the CNN in generating adaptive γ parameters is also presented.

6.2.2.1 Quality

The method of obtaining quality results for PTF_p is presented in Figure 6.4. The method used for the PTF_a results is presented in Chapter 5 and for PTF_4, PQ and HLG

	Display-Referred		Scene-Referred	
Method	**puPSNR**	**VDP**	**puPSNR**	**VDP**
PQ	56.27	74.86	—	—
HLG	56.91	79.51	51.43	75.94
PTF_4	57.83	78.54	56.05	73.06
PTF_a	59.43	81.78	57.15	75.14
PTF_p	59.32	81.67	56.57	75.92

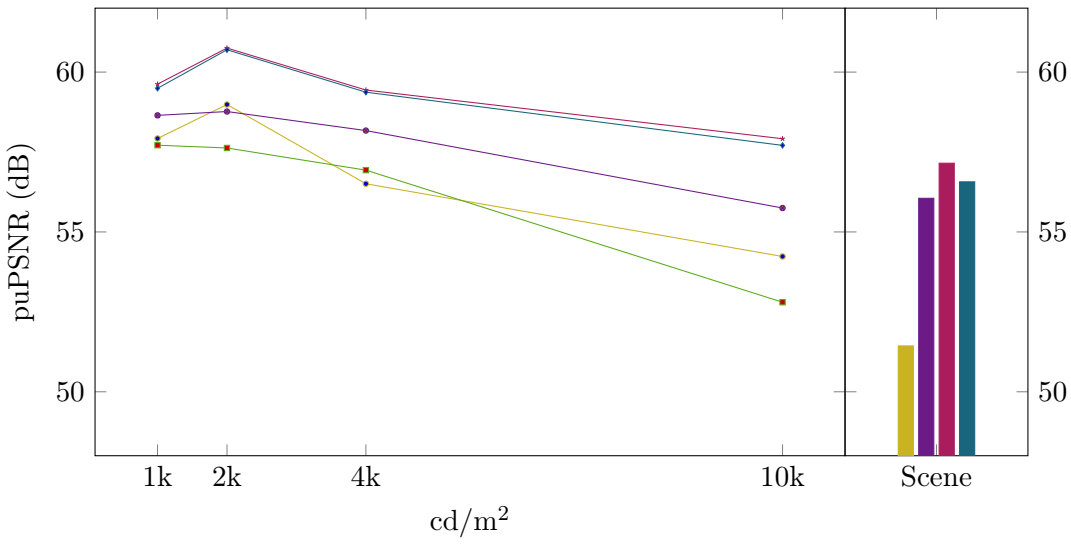
Table 6.1: Scene-Referred and Mean Display-Referred coding performance in (a) HDR-VDP-2.2.1 and (b) puPSNR for PQ, HLG, PTF_4, PTF_a, and PTF_p when tested on 74 images selected from a dataset of 735 images. The results for PQ on scene-referred data are not reported.

in Chapter 3. The parameters used in this evaluation are the same as those used for Chapter 5. The PTF compression function produced planar 10-bit $Y'C_BC_R$ frames with a 4:2:0 chroma-subsampling. Base-band coding only results were produced as the CNN was trained on the coding-only results presented in Chapter 5 and as Section 5.2 has demonstrated, γ selection is dependent on QP. The network would need to be trained with QP as an input to predict the ideal γ value for compressed encoding, however many factors, such as GOP structure, would affect the model.

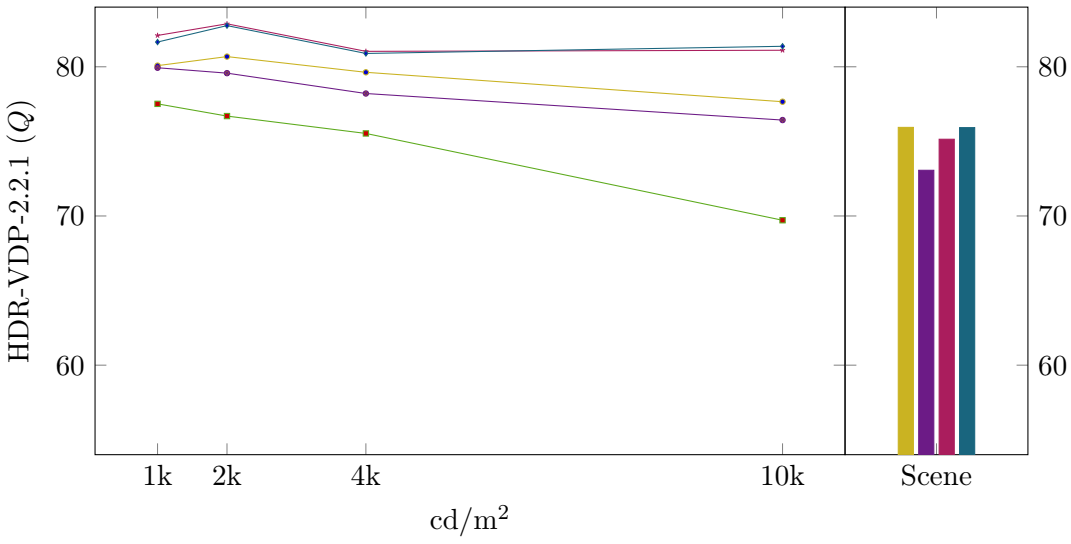
Figure 6.5 presents the coding performance of PTF_p against the other tested methods and using the metric introduced earlier. For both metrics, the results of PTF_p closely match those of PTF_a demonstrating that the CNN is predicting high performing γ values from the images. The average loss when compared with PTF_a is 0.11 puPSNR and 0.11 VDP for display-referred data, and 0.58 puPSNR for scene-referred data. PTF_p performed better than PTF_a for scene-referred data when tested with HDR-VDP-2, gaining 0.78 VDP. PTF_p performs better than PQ, HLG and PTF_4 for display-referred data. For scene-referred data, PTF_p performs better than HLG and PTF_4 for HDR-VDP-2, and better than PTF_4 for puPSNR but 0.02 VDP lower than HLG. PQ results are not presented for scene-referred data as the method cannot represent values $> 10{,}000\,\mathrm{cd/m^2}$. Table 6.1 presents precise numbers for the methods and metrics in a tabular format.

6.2.2.2 Performance

The feed-forward performance of the network is important as that determines how quickly the network is able to generate adaptive γ parameters. The aim is to enable real-time adaptive γ prediction, where real-time is classed as a rate in excess of 60 fps. The network computations were executed on a Nvidia Geforce GTX1080 GPU using PyTorch.



(a) puPSNR



(b) HDR-VDP-2

Figure 6.5: Coding performance in (a) HDR-VDP-2.2.1 and (b) puPSNR for PQ, HLG, PTF_4, PTF_a, and PTF_p when tested on 74 images selected from a dataset of 735 images. The results for PQ on scene-referred data are not reported as the method is limited to representing data of 10,000 $\mathrm{cd/m^2}$ or less.

The testing fold was used for the performance analysis and contained 3-channel, 32-bit floating-point RGB images at a resolution of 1920×1080. The network was capable of predicting adaptive parameters in 15 ms discounting the time taken to load the images into memory. This equates to 66.6 fps therefore achieving the aim. It should also be noted that the network was not tuned for performance and further performance gains can be expected as the network and surrounding processes are optimised. It is expected that implementing the network in a C++ environment rather than PyTorch would increase performance.

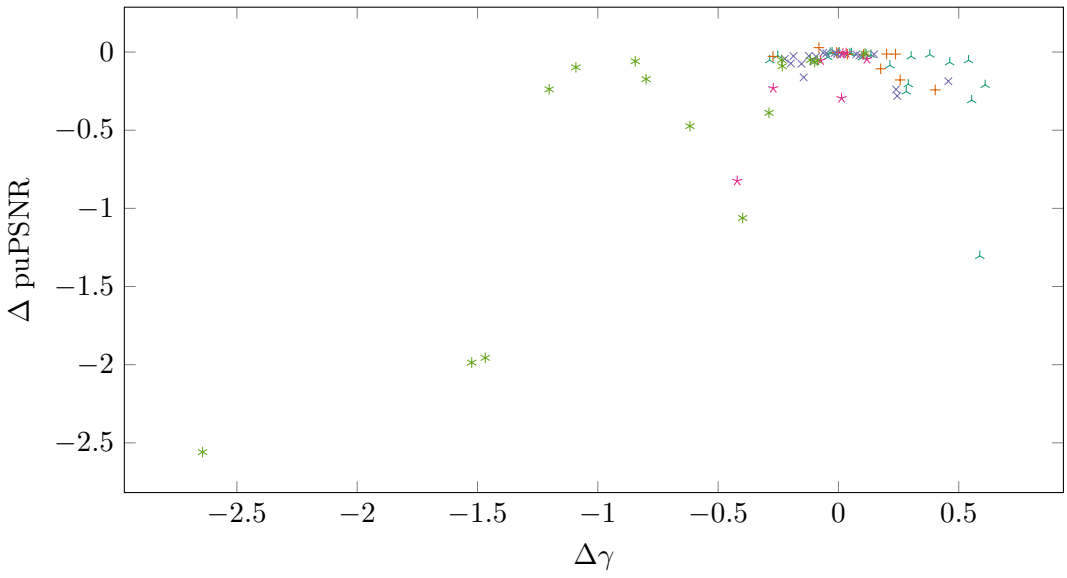
6.2.3 Analysis & Discussion

The results show that using a CNN can accurately predict the γ values required for high-quality compression. The mean absolute error of the network after training was 0.303 however this loss has not significantly impacted the puPSNR and HDR-VDP-2 results produced when using predicted γ values. One hypothesis as to why the relatively high error has not impacted performance is that some images are impacted much more heavily by γ variation than others. If the error in prediction occurs only in images where there is a broad range of high-quality parameters, then the quality metrics will report a similar result. Another factor may be that although predicting a γ beneath the optimal quickly causes a loss in quality, predicting a γ above the optimal does not cause such a loss. Both of these factors can be seen in Figure 5.2.

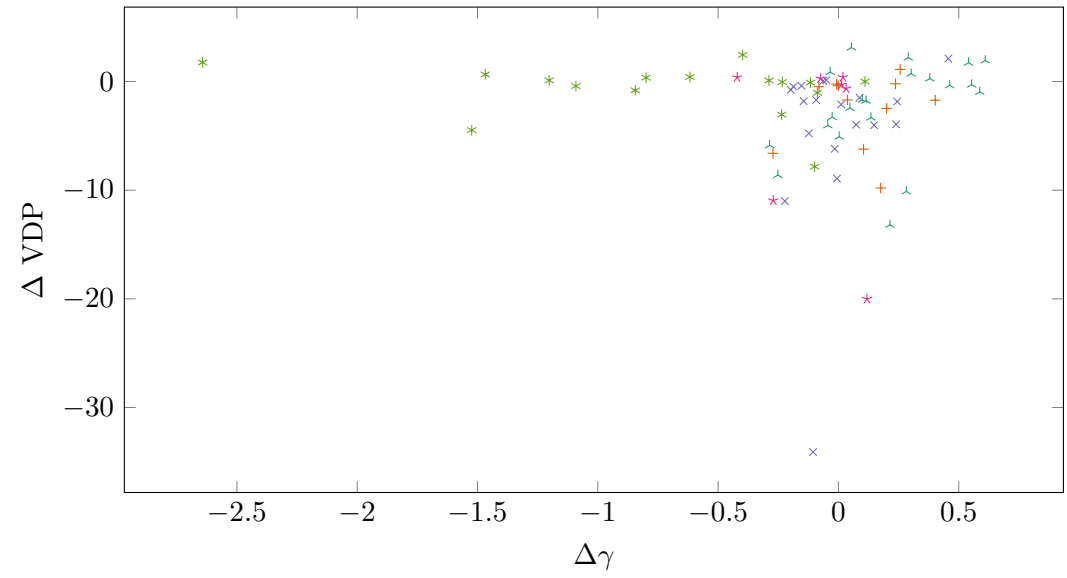
The difference in γ and metric result between the optimal dataset produced by PTF_a and the predicted data used in PTF_p is presented in Figure 6.6. As puPSNR was the metric used to train PTF_a, other than a single outlier, a close relationship can be seen between a mis-prediction of γ, $\Delta\gamma$, and a change in puPSNR, ΔpuPSNR. The relationship between $\Delta\gamma$ and ΔVDP is less distinct. In fact, in some cases a significant change in γ results in a higher HDR-VDP-2 result. It should be noted that the mis-prediction of γ is more extreme when the data is scene-referred, and the resulting cost in terms of puPSNR is pronounced. Also, a significant loss of quality in one metric does not necessarily relate to a loss in another metric. A good example of this is the data point $\Delta\text{VDP} = -34.1$, the mis-prediction in $\gamma = -0.106$, however the change in puPSNR is only -0.056. The cause of this effect is not known as both metrics correlate well with human perception [MDB+16a].

6.3 Conclusion & Future Work

The results show that using a CNN to predict γ values used in the compression of HDR imagery by PTF_p can result in high-quality compression. The error between the testing



(a) puPSNR



(b) HDR-VDP-2

Figure 6.6: Analysis of how γ mis-prediction affects quality between the optimised dataset and the predicted dataset. (a) shows how a change in γ results in a change in puPSNR. (b) shows how a change in γ results in a change in VDP. Each point is colour-coded to show which luminance level the result is from.

dataset and the predicted γ values was 0.303. The loss in quality introduced when predicting γ values from image content rather than optimising for compression was, for the dataset used in this evaluation, 0.284% puPSNR for display-referred data, 1.02% puPSNR for scene-referred data, 0.133% VDP for display-referred data and a gain of 1.04% VDP for scene-referred data. These losses are slight and PTF_p outperforms the other tested methods, PTF_4, HLG and PQ for display-referred data across both metrics, outperforms PTF_4 on scene-referred data for both metrics and outperforms HLG on scene-referred data for VDP. The computational speed up is in the region of four orders of magnitude and sufficient to predict adaptive γ values in a real-time environment, however there is an associated cost in terms of memory requirements.

This chapter has also shown that a regression model built from simple image statistics does not have the ability to predict the γ values needed for high-quality compression. This is evidenced by the high error between the testing dataset and the predicted γ values of 1.06. A more complex machine learning method such as an ANN might be able to learn a more accurate model using simple image statistics however it then incurs the memory cost. Other potential avenues would be to include more statistics into the dataset such as a histogram or an investigation into the features learned by the CNN to create a more accurate regression model. These will be the topic of future work. Another topic will be to include more information to the fully-connected stage of the CNN such as the QP. This will allow the network to predict the ideal γ required for specific quality settings instead of simply for base-band coding.

The results in this chapter show that machine learning, and specifically a CNN, can be highly-accurate in predicting the γ parameters needed for high-quality compression. This chapter introduced a new compression method based on a CNN, PTF_p, which outperforms other proposed methods in terms of quality for HDR video compression. The γ values predicted by the proposed CNN result in near-optimal compression quality.

Chapter 7

Conclusions

This thesis has introduced an efficient and adaptable approach to HDR video compression. The proposed method enables content, display and environmental adaption for HDR video at a compression rate necessary for real-time encoding.

Chapter 3 introduced PTF, a novel approach to HDR video compression which an evaluation concluded provides high-quality compression using computationally efficient techniques. Chapter 4 reported a subjective evaluation which demonstrated the need for compression capable of adapting to both the display and the ambient environment. Chapter 5 developed a framework into which PTF can be extended to provide compression that adapts to the display and the ambient environment via the content. Finally, Chapter 6 optimised the performance of content adaption by developing a deep learning model to predict the adaption required for optimal compression performance based on the content. This chapter presents a summary of the contributions presented by this thesis and suggests directions for future work.

7.1 Efficient HDR Video Compression

An efficient, high-quality HDR video compression method was proposed in Chapter 3. It provides improved objective quality when compared with other state-of-the-art methods at a reduced computational cost. The method, PTF_4, applies traditional power functions to create high-quality compressed HDR video. The chapter presents a comprehensive objective evaluation showing that the proposed method reduces the bit-rate required to maintain quality by 4.8% over PQ and 33.2% over HLG, two leading methods in the field.

Furthermore, power functions are computationally efficient, with a power-of-two parameter requiring only a chain of products to decompress. The proposed method only requires two multiplications per value to decompress HDR video. The method

is shown to decompress HDR video frames at over 380 fps in a single-threaded CPU implementation, 29.5 times faster than another state-of-the-art method and 1.5 times faster than that achieved by using a look-up table without the associated memory requirement. Due to its straightforward nature, this method is amiable to hardware acceleration via a GPU or FPGA implementation.

The research question posed at the start of the chapter was answered by demonstrating that high-quality video compression can be achieved with a straightforward transfer function.

The proposed method contains a single parameter for which a range of values were tested. The results recommend the integer value 4, as it provides high performance and the best overall quality when tested with a range of sequences. The evaluation suggests that a fixed transfer function might not result in the highest possible quality. It was noted that this value could be adjusted and that some of the tested sequences suggest other values might result in higher-quality compression.

7.2 Effects of Ambient Illumination and Peak Luminance

Chapter 3 made the assumption that the luminance of the content was fixed, whereas Chapter 4 hypothesised that the necessary luminance might be a condition of the ambient environment at the display. A subjective evaluation was performed involving 40 participants over a range of ambient illuminations and display luminances to ascertain the dynamic range they were able to perceive. The environments tested were chosen to reflect the locations where HDR displays might be used and ranged from the near-black illumination of a theatre to the bright sunlight of an outdoor display. The participants were asked to report the detail visible on three pre-prepared test charts each containing a different task. For each group of participants, the experiment was performed at a fixed ambient illumination but at a range of different display and test chart luminance levels to understand how ambient illuminance affects what is perceivable.

The main effect of the peak luminance of the display was found to be significant ($p < 0.01$) across all three conditions and likewise, the main effect of the ambient illumination was found to be significant ($p < 0.01$) across all three conditions. This evaluation concluded that the dynamic range that can be perceived on a display is dependent on the peak luminance of the screen and the ambient illumination of the environment. From this, the required luminance for content of a given dynamic range can be calculated from the ambient illumination. For the detail to be perceived, the luminance of the content requires adaption based on the ambient environment at the display.

7.3 Adaptive Compression for Content, Display and Environment

Chapter 5 demonstrated the benefits of adaptive content on HDR compression performance and proposed an extension to PTF which increases performance by optimising the compression curve. It showed that as the luminance of the content is adjusted, the optimal compression curve adapts along with it. However, it was also shown that a single compression curve was not ideal for all content, and that adapting the curve to the specific content could yield improvements. A framework was proposed that allows features of the content, display and environment to be taken into account when choosing the method of HDR compression. The method proposed in Chapter 3 was extended to form PTF_a, which utilises the framework to provide compression that adapts to the luminance of the display and the content to provide improved quality. (The environment is accounted for by adjusting the luminance of the content to ensure the detail is visible.) The method utilises a function minimisation algorithm to search for the optimal compression curve. The implementation proposed minimised the reciprocal of the puPSNR metric.

The proposed extension was evaluated on a dataset of 147 images from a range of sources, graded to 4 display-referred luminance levels and the scene-referred luminance. The adaptive method was evaluated against PTF_4, PTF_8, PQ and HLG using puPSNR and HDR-VDP-2. The results showed that PTF_a exhibited improved objective quality over the tested methods at all luminance levels and across both metrics. The puPSNR result was 2.2% higher than PTF_4 and the HDR-VDP-2 result was 3.5% higher for display-referred data. This demonstrated the robust nature of the proposed method as it did not overfit the compression curve to puPSNR and exhibited improved quality when tested with HDR-VDP-2. This chapter showed that adapting the compression method to the content, ambient environment and luminance of the display improves compression.

The evaluation also revealed some other interesting aspects such as the propensity of the compression curve to become more aggressive as the luminance of the content increased, likewise the compression curve became more aggressive as the content was encoded at a lower bit-rate. Both of these effects demonstrate the need for adaptive compression.

Chapter 5 suggested solutions to the performance penalty incurred by using an iterative optimisation method, and an investigation of one of these solutions formed the basis of Chapter 6.

7.4 Deep Learning for Adaptive Compression

Chapter 5 verified the improvements adaptive compression provides, while Chapter 6 made the technique practically applicable. The performance of display, environment and content adaptive PTF was increased to that necessary for operation in a real-time solution. Both regression using simple image statistics and deep learning models were evaluated, however regression did not improve on the performance of a fixed compression curve.

A convolution neural network (CNN) was trained with deep learning to predict the curve for optimal compression when presented with content. The training dataset was that produced by the evaluation in Chapter 5 which had calculated the optimal compression curves. A CNN was employed as it allowed the features that are important in predicting the optimal compression curve to be learned by the network, rather than provided by pre-processing in the case of regression. The CNN also resulted in more accurate predictions. The error in the predictions did not cause a particular drop in objective quality as the evaluation showed. PTF_p, the CNN based method, generally experienced a slight loss in quality when compared with the optimal results from PTF_a, however in one instance it showed improved quality over PTF_a. Overall, the method resulted in a higher objective quality than the remaining tested methods.

The benefits of adaption were made broadly applicable with the increase in computational performance afforded by the prediction of compression parameters by a CNN. In feed-forward operation, the CNN could predict curve parameters from the content to provide high-quality compression in 15 ms for 1920×1080 HDR images. The CNN has increased the performance of adaptive compression by four orders of magnitude with little cost in compression quality.

7.5 Contributions

This thesis set out to answer the following research question:

> *Can the quality of computationally efficient High Dynamic Range video compression be improved by adapting to the environment and display?*

The primary contribution of the thesis is an answer to this overarching question. The work has shown that adaptive HDR video compression is possible in a computationally efficient manner using the methods and techniques developed. The further contributions of this thesis are as follows:

1. A novel HDR video compression method providing improved objective quality at reduced computational cost (Chapter 3).

2. A subjective evaluation involving 40 participants to demonstrate the impact of ambient illumination of the detail perceived on an HDR display (Chapter 4). It showed that for HDR to be perceived in bright ambient conditions, a display must have a high peak luminance.

3. A compression method that adapts to the content, display luminance and the ambient environment at the display to provide the optimal compression (Chapter 5). The method demonstrated improved quality across multiple metrics showing the proposed method is a robust form of adaption.

4. An adaptive compression method that uses deep learning to predict the ideal compression curve from the content, display luminance and ambient environment (Chapter 6). The CNN was capable of predicting compression curves at over 60 fps, four orders of magnitude faster than the full optimisation, enabling content adaptive compression at real-time frame rates.

7.6 Limitations

The research in this thesis is only concerned with the investigation of power functions due to the inherent ease of implementation and straightforward numerical analysis they provide. It is probable that other classes of function could be employed to similar or greater effect. This thesis has shown that power functions provide high quality compression with minimal computational cost. However, the straight line that they produce through contrast sensitivity plots fits the human visual system only to a point, as they provide unnecessary detail in the darker regions of the image. The techniques of adaption that were developed efficiently using PTF could be employed on other functions to produce similar benefits.

This thesis demonstrated that the proposed predictive model produces high quality compression, it did not extend to evaluating the model using video footage. While predicted γ parameters could be applied individually to each video frame, this may introduce flickering which would reduce the compression efficiency. Temporally smoothing the predicted parameter could reduce this effect. Alternatively, this could be achieved by using a single predicted parameter for each scene or group of pictures (GOP).

7.7 Future Work

The work presented in this thesis has raised a number of further questions and directions for study. This section will discuss those avenues.

7.7.1 Subjective evaluation

Although the objective metrics used to evaluate the methods proposed in this thesis have been shown to correlate highly with subjective tests [HBP+15; MDB+16a], a subjective evaluation of the proposed methods (PTF, PTF_a, PTF_p) would provide more evidence regarding the quality of the solutions proposed and further improve the confidence in the objective results.

Adaptive compression promises to ensure that the range and detail in the content is visible irrespective of the ambient illumination. A subjective experiment to test this hypothesis would be beneficial in promoting HDR content in a wider range of environments. This evaluation would also provide valuable data on the performance of compression curves under ambient illumination and test the hypothesis that the adaptive technique proposed in Chapters 5 and 6 improve quality.

7.7.2 Extended adaption

While the adaptive method proposed in Chapter 5 takes into account the ambient illumination via the peak luminance of the content and the content itself, it does not take into account properties of the encoding environment such as the bit-depth, chroma sub-sampling, the encoded bit-rate or quality parameter, or the encoder itself. The chapter showed that the quality parameter has an effect on the ideal compression curve, so it is expected that other encoder settings would have an effect on the ideal curve also. Chapter 3 demonstrated that bit-depth did not appear to affect the ideal compression curve, however in light of the work performed in later chapters this should be investigated further.

The adaption proposed took a single input, the content, and produced a single output. Increasing the adaption to all of the proposed parameters would greatly increase the time required to calculate the optimal parameters. However, with the model proposed, it would be a one time cost which could be used as an input to the CNN. The model may then be able to predict the appropriate curve given all of the parameters.

7.7.3 Improved regression

Regression techniques and artificial neural networks (ANNs) both offer computational performance benefits over CNNs. If a model could be developed using either of these methods, the computational performance of adaptive compression could be increased further, potentially to the point of enabling the technique to operate in hardware. One potential strategy that could be used to improve the quality of either regression or an ANN would be to extract the image features that the CNN has learnt to calculate them

explicitly.

7.8 Final Remarks

During the course of this study, HDR video compression has moved from an academic research area to an industry standard. However, outstanding problems remain. An evaluation has shown that for HDR to be perceivable in many common usage scenarios, a higher peak luminance is needed than is typically available via commodity displays. The proposed PTF_a adaptive compression method has been designed to provide optimal compression in variable ambient conditions where other methods may be impacted. PTF_p has accelerated the adaptive method such that it can be used in a real-time system, enabling the use of adaptive compression in live broadcast environments. Furthermore, the work in the thesis provides information on how compression methods respond to changing environments and the adjustments that are needed to maintain the highest possible quality. This thesis provides a step towards taking HDR out of a dark room and in to the living room.

Appendix A Warwick Dataset

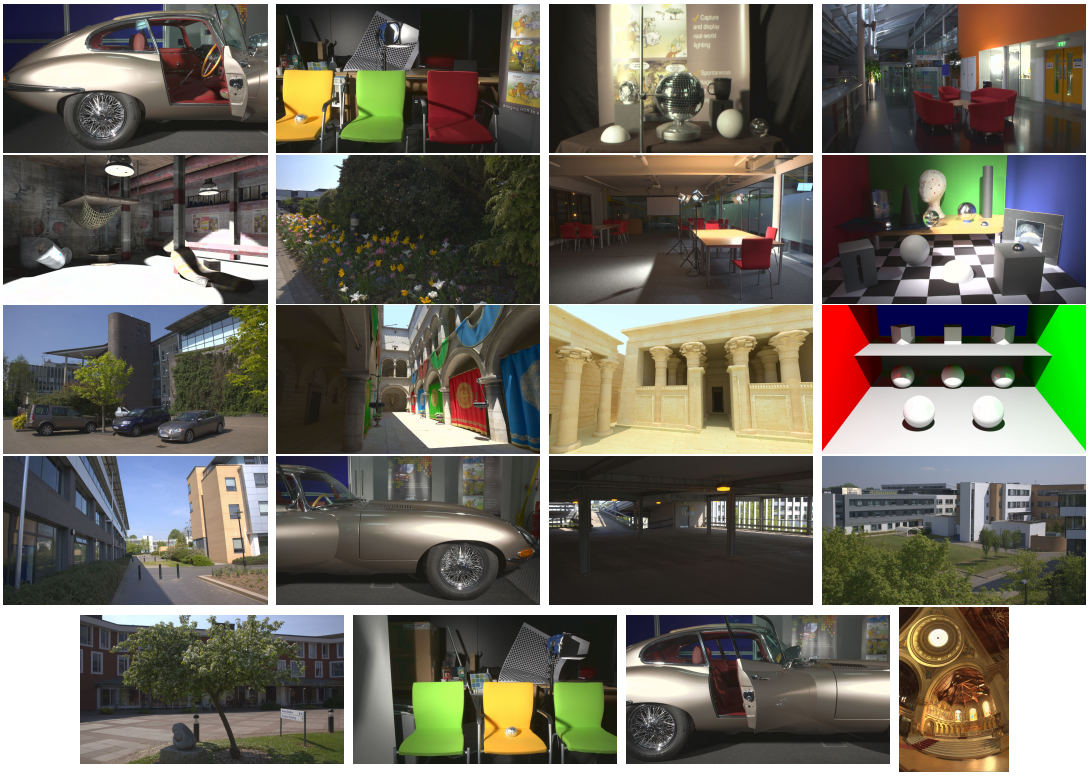
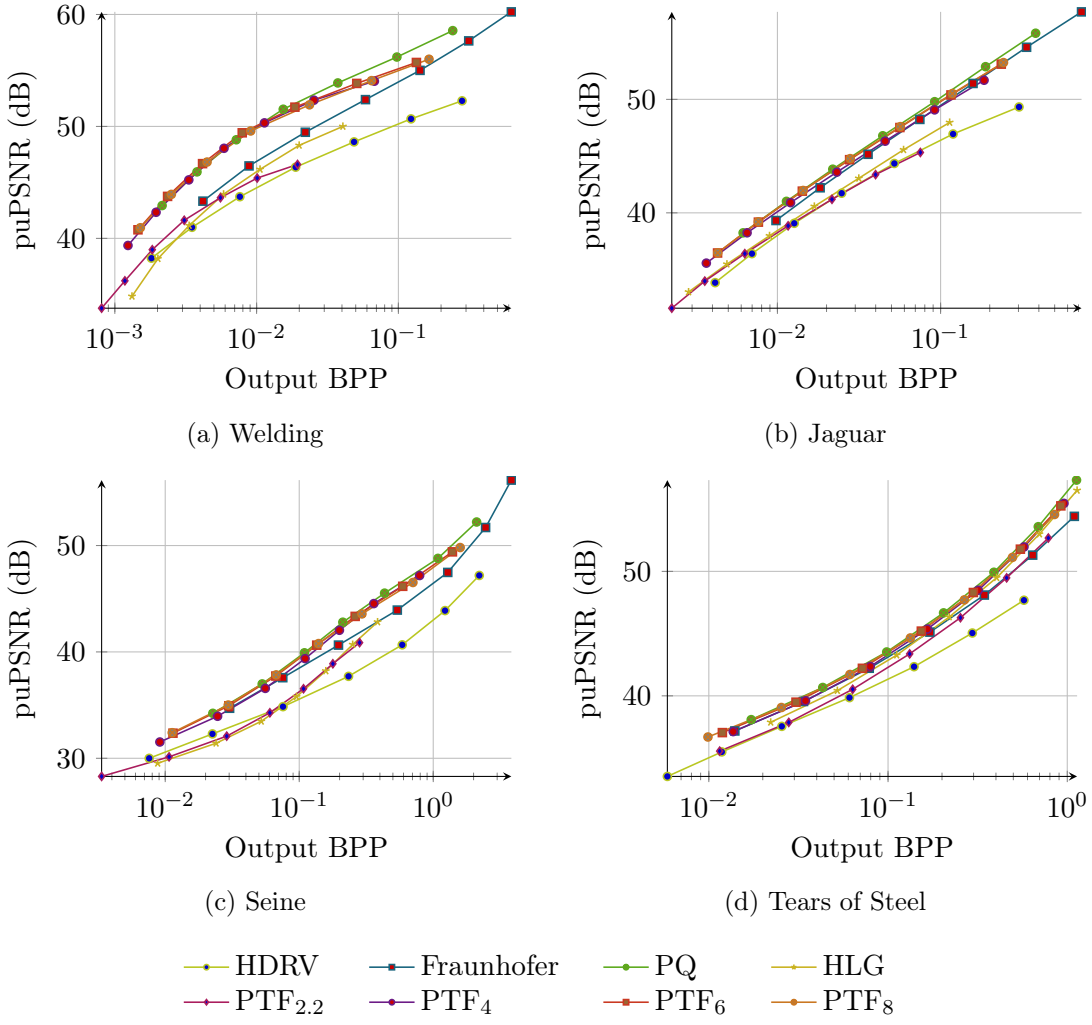


Figure A.1: Images used for evaluation of γ variation at a selection of bit-depths.

Appendix B PTF Results

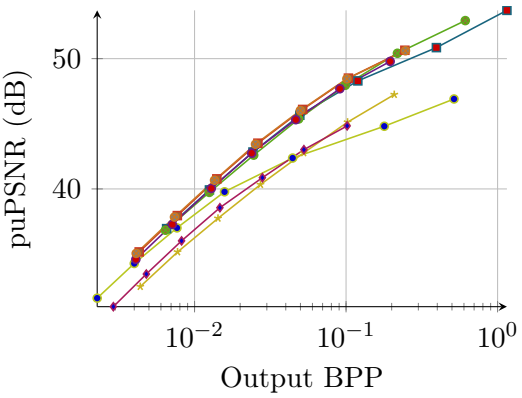


(a) Welding

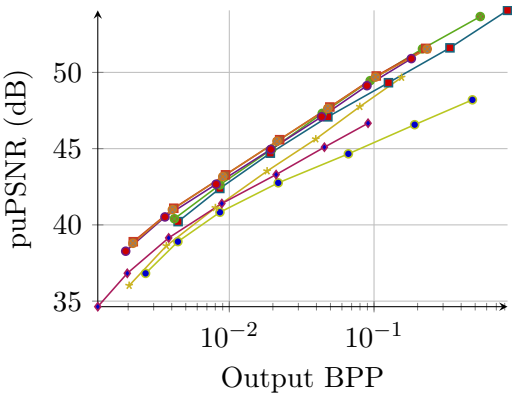
(b) Jaguar

(c) Seine

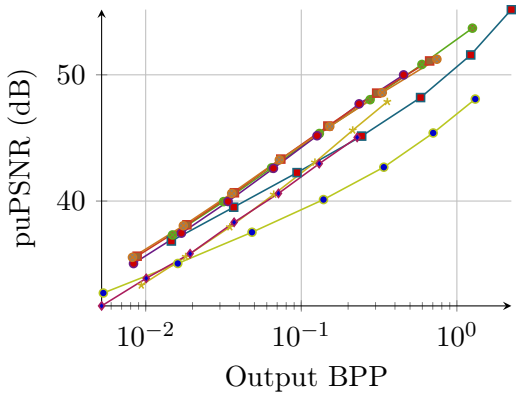
(d) Tears of Steel



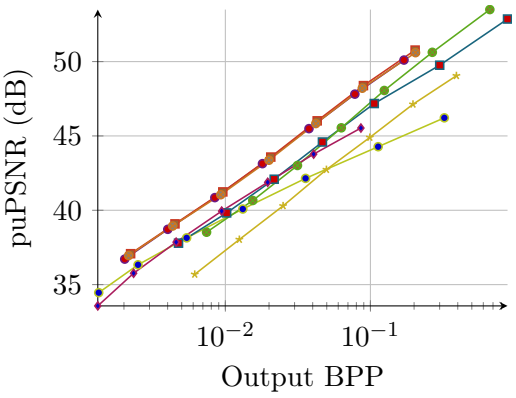
(e) Mercedes



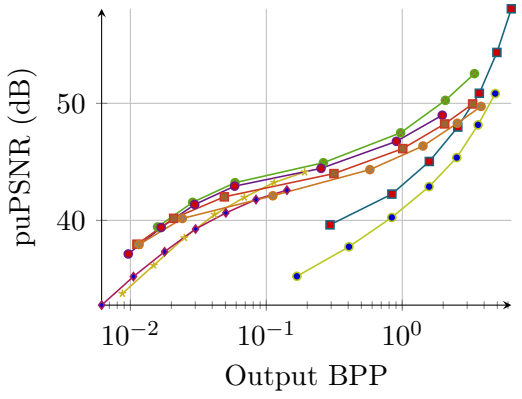
(f) Beer Festival 4



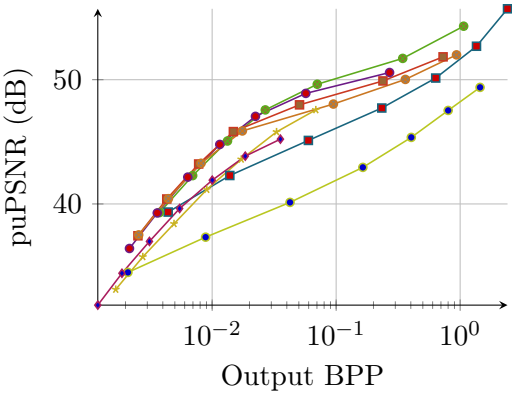
(g) Carousel Fireworks 9



(h) Bistro 3

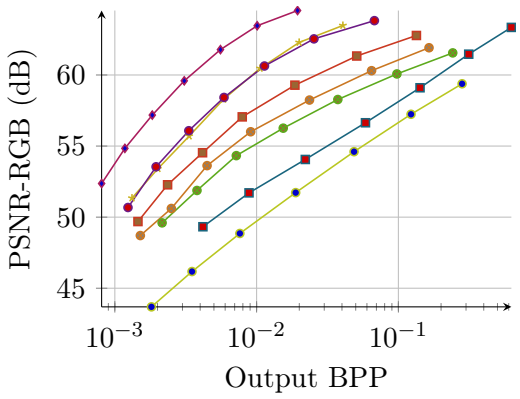


(i) Fireplace 2

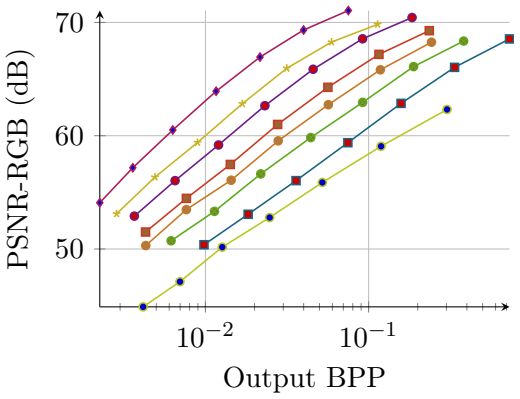


(j) Showgirl 1

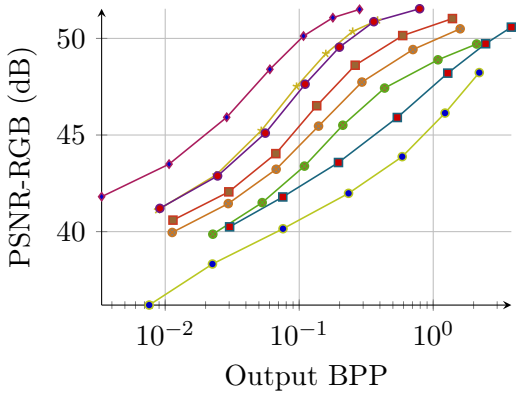
Figure B.1: Rate distortion characteristics showing how the different HDR video compression methods perform on a variety of sequences. The rate is measured in output bits per pixel (BPP) and the distortion as a puPSNR. Figures are presented with a logarithmic x axis to improve clarity.



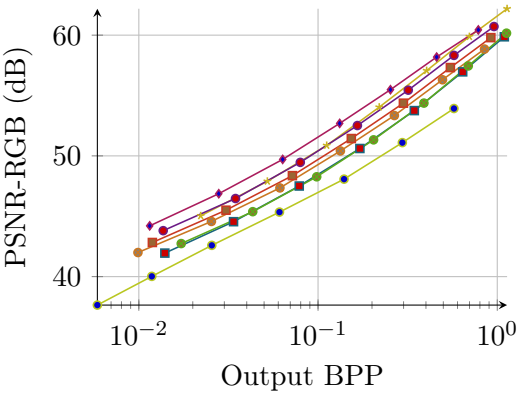
(a) Welding



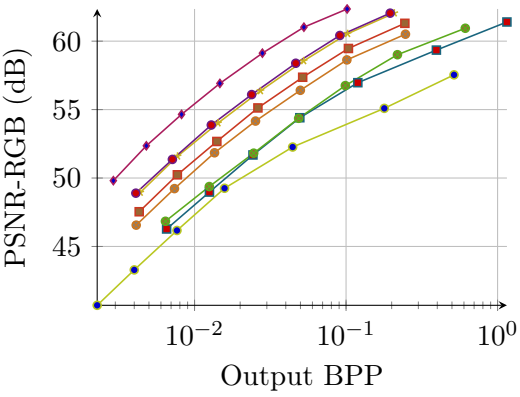
(b) Jaguar



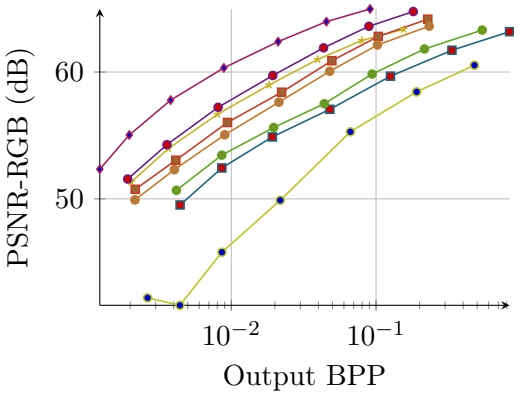
(c) Seine



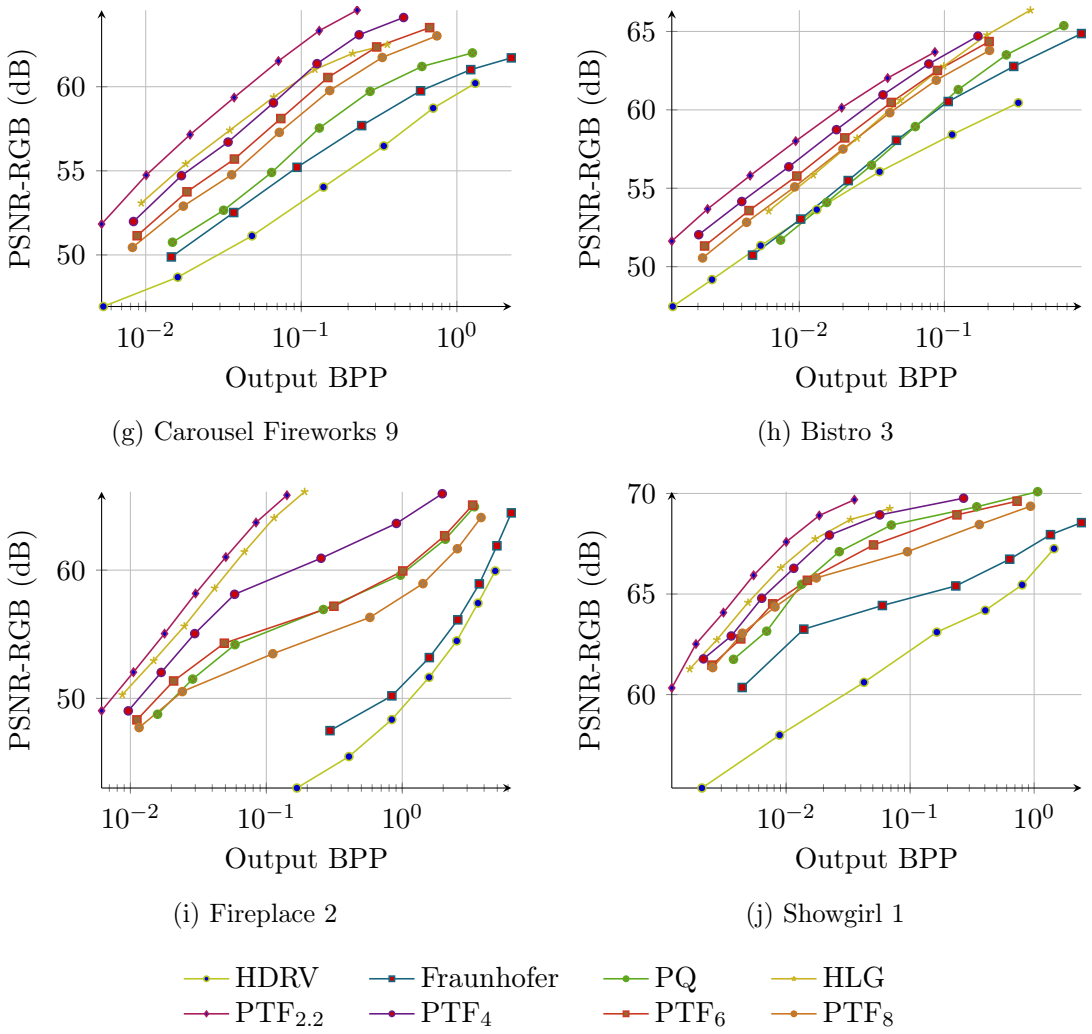
(d) Tears of Steel



(e) Mercedes



(f) Beer Festival 4



(g) Carousel Fireworks 9

(h) Bistro 3

(i) Fireplace 2

(j) Showgirl 1

Figure B.2: Rate distortion characteristics showing how the different HDR video compression methods perform on a variety of sequences. The rate is measured in output bits per pixel (BPP) and the distortion as a PSNR. Figures are presented with a logarithmic x axis to improve clarity.

Appendix C Ethical Approval for Ambient Illumination Experiment

The data collection for the experiment presented in Chapter 4 was conducted by INESC TEC at Universidade de Trás-os-Montes e Alto Douro (UTAD), Vila Real, Portugal. This appendix contains a letter from Dr Maximino Bessa, the head of the laboratory, confirming that the relevant ethical approval was obtained for the experiment. Subsequent ethical approval for the transfer of the collected data from Portugal to Warwick and further analysis was sought and granted from the University of Warwick. This approval, BSREC REGO-2018-2184, is included in this appendix. UTAD was chosen as the venue for the experiment as the SIM2 display was available, and bright sunshine was easily obtained for the required ambient light levels.

INESCTEC

To whom it may concern

This letter is to confirm all the appropriate ethical approval required by INESC TEC were in place for the user studies conducted at INESCTEC - MASSIVE Laboratory at Universidade de Trás-os-Montes e Alto Douro1 (UTAD) in Vila Real, Portugal in May 2017.

As no personal data of the participants was collected as part of this user study, INESC TEC does not require a formal ethical approval application, for this type of study.

If you require further information, please just let me know.

Yours sincerely

[redacted]

Dr Maximino Bessa

INESCTEC

INESC TEC
Instituto de Engenharia de Sistemas e Computadores, Tecnologia e Ciência
Associação privada sem fins lucrativos declarada de utilidade pública
Pessoa Coletiva 504 441 361 • CRC Porto

Campus da FEUP
Rua Dr. Roberto Frias
4200 - 465 Porto
Portugal

T +351 222 094 000
F +351 222 094 050
info@inesctec.pt
www.inesctec.pt

Biomedical and Scientific Research Ethics Committee (BSREC):

Application Form for Research Ethical Approval

SECTION 1. APPLICANT DETAILS

1.1 RESEARCHER

Researcher's Title: Mr
Researcher's Forename: Jonathan
Researcher's Surname: Hatchett

Researcher's Faculty/School and Department: WMG

Researcher's Status:

Undergraduate Student	☐
Taught Postgraduate Student	☐
Research Postgraduate Student	☒
Staff	☐
Other	☐

Please specify:

If Student:
Name of course/qualification: PhD in Engineering

If Staff:
Researcher's Post:

1.2 RESEARCHER'S CONTACT DETAILS

Warwick e-mail address: j.p.hatchett@warwick.ac.uk
Daytime telephone number: 07745513848

1.3 SUPERVISOR (COMPLETE FOR ALL STUDENT PROJECTS)

Supervisor's Title:	Prof
Supervisor's Forename:	Alan
Supervisor's Surname:	Chalmers
Supervisor's Post:	Professor
Supervisor's Faculty/School and Department:	WMG
Supervisor's Warwick e-mail address:	Alan.Chalmers@warwick.ac.uk
Supervisor's daytime telephone number:	024 765 22102

SECTION 2. PROJECT DETAILS	
2.1 Project Title:	Psychophysical Evaluation of Perceptible Detail in Natural Environments
2.2 Estimated Start Date of Project:	10/4/2018
2.3 Estimated Completion Date of Project:	10/6/2018
2.4 Sponsoring Organisation: (for University of Warwick staff and students, undertaking non-commercial projects, this will be the University of Warwick)	The University of Warwick
2.5 Funder: (e.g. unfunded student project, unfunded Departmental project, Medical Research Council (MRC), Economic & Social Research Council (ESRC), EU)	EPSRC

2.6 TYPE OF PROJECT

Is the project:

Primary Research	☐
Research limited to the use of previously collected identifiable data	☐
Research limited to the use of previously collected anonymised data	☒
Clinical Audit	☐
Service evaluation or Development	☐

Other- please specify:

2.7 LINKS WITH OTHER BSREC APPLICATIONS

Is the project linked to any other BSREC application? No
If yes, detail:
Project title:
Chief Investigator:
BSREC Reference (if known):

Nature of linkage:

2.8 LOCATION

Will any part of the project be undertaken overseas? Yes

State all of the locations at which the project will be undertaken, whether in the UK or overseas e.g. public place, school (a), school (b) etc., in researcher's office:

Visualisation Group Office, WMG, University of Warwick

Department of Computer Science, UTAD, Vila Real, Portugal

2.9 PARTICIPANTS

State the total number of planned participants: 40

BREAKDOWN OF PARTICIPANTS

Where applicable, state the breakdown of participants by type and number of each type of participant, e.g. children, parents, teachers, etc.:

Type of Participant:	**Number:**
Ambient Environment Experiment Participant	40

SECTION 3. TRAINING	Yes	No
Have you successfully completed the Research Integrity online training module? https://www2.warwick.ac.uk/services/ldc/researchers/opportunities/development_support/research_integrity/	☐	☒

SECTION 4. RISK AND ETHICAL CONSIDERATIONS CHECKLIST

Complete the checklist ticking '**Yes**' or '**No**' to all questions.
Note that, where you have ticked '**Yes**' to a question below, you will need to specifically address the ethical issues raised by that point in the study protocol.

		Yes	No
A	Does the study involve **participants who are particularly vulnerable** or **unable to give informed consent** or **in a dependent position** (e.g. children, your own students, over-researched groups, people with learning difficulties, people with mental health problems, young offenders, people in care facilities, prisoners)?	☐	☒

B	Will participants be taking part in the study without their consent or knowledge at the time, or will **deception** of any sort be involved (e.g. covert observation of people in non-public places)?	☐	☒
C	Is there a risk that the highly sensitive nature of the subject might lead to **disclosures** from the participant concerning their involvement in illegal activities or other activities that represent a threat to themselves or others (e.g. sexual activity, drug use, or professional misconduct)?	☐	☒
D	Could the study induce **psychological distress or anxiety**, or produce **humiliation**, or **cause harm**, or lead to **negative consequences** beyond the risks encountered in normal life?	☐	☒
E	Does the study involve **substantial physical exertion**?	☐	☒
F	Does the study involve the **administration** of any substance?	☐	☒
G	Does the study involve **physically intrusive procedures**, use of **bodily materials** or **human tissue**, or **DNA/RNA analysis**?	☐	☒
H	Is any **reward**, apart from travelling and other expenses, to be given to participants?	☐	☒
I	Does the study involve **collaboration with any company** or **organisation** external to the University of Warwick?	☒	☐
J	Could the proposal give rise to researchers having any **conflicts of interest**?	☐	☒
K	Does the study involve the **NHS**? (staff, anonymous data or facilities)*	☐	☒
L	Will the researchers go to any areas where their **safety may be compromised**?	☐	☒
M	Will **pregnant women** be participants in the study?	☐	☒
N	Will the study involve children **under 5 years** old?	☐	☒
O	Is the research commissioned by the **military**?**	☐	☒
P	Is the research commissioned under an **EU security call**?*	☐	☒
Q	Does the research involve the acquisition of **security clearances?***	☐	☒
R	Does the research concern **terrorist or extreme groups**?*	☐	☒

** Please note studies involving **NHS** patients, their **identifiable** data or **tissue** require review by an NHS REC (not BSREC)*

*** Please refer to the University webpages on Prevent Duty*

SECTION 5. SIGNATURES AND DECLARATIONS

5.1 RESEARCHER/APPLICANT

I undertake to abide by the University of Warwick's Research Code of Practice in undertaking this study.

I understand that BSREC grants ethical approval for projects, and that the seeking and obtaining of all other necessary approvals and permissions prior to starting the project is my responsibility.

I understand that I must not begin research and related projects with human participants until I have received full approval from the relevant Research Ethics Committee of the University of Warwick.

I understand that any changes that I would like to make to this study after receiving approval from BSREC must follow BSREC procedures as detailed on the BSREC web pages.

Name of Researcher: Jonathan Hatchett

Signature: ……
Date: 09/03/18

Send a signed copy of the form to bsrec@warwick.ac.uk, along with copies of **all** study documentation, including questionnaires, interviews schedules/topic guides, posters/leaflets, invitation emails etc.

5.2 SUPERVISOR SECTION

I confirm that I have read this application and will be acting as the student researcher's supervisor for this project.

The proposal is viable and the student has the appropriate skills to undertake the research. Participant recruitment procedures, including the Information Leaflet(s) to be provided and the process for obtaining informed consent, are appropriate, and the ethical issues arising from the project have been addressed in the protocol.

I have reviewed any questionnaires, interview schedules/topic guides where relevant, and these are appropriate for the project.

I understand that BSREC grants ethical approval for projects, and that the seeking and obtaining of all other necessary approvals and permissions prior to starting the project is the responsibility of the student.

I understand that research and related projects with human participants must not commence without full approval from the relevant research ethics committee of the University of Warwick.

Name of Supervisor: Alan Chalmers

Signature: ……
Date:

NB: An e-mail from the Academic Supervisor that states the above, in lieu of a signature on this form, may be sent to: bsrec@warwick.ac.uk

Proposal: Data Analysis of Perceptible Detail in Natural Environments

Jonathan Hatchett

10th April 2018

1 Summary

This document presents a proposal for the secondary analysis of previously collected data for publication. The data involves an evaluation performed by human participants who were asked to report the amount of detail they could perceive on a display at a range of brightnesses and ambient illuminations.

This evaluation was performed at Universidade de Trás-os-Montes e Alto Douro[1] (UTAD) in Vila Real, Portugal in May 2017 and was covered by ethical approval granted to INESC TEC[2]. The consent form was produced in Portuguese and is included in Section 8. This application is for the data sharing required to analyse the results by members of the University of Warwick working on collaborative publications with academics at UTAD.

The data to be transferred will contain the ages, genders and responses of each participant. The ages and genders will only be used to ensure a broad cross section of the population is sampled.

2 Background

High Dynamic Range (HDR) imaging has the potential to capture, store, process, transmit, and display the full range of light in a scene [3]. No stage in the image processing pipeline will require information be discarded, which typically manifests as under- or over-exposed pixels in the image. The information provided by HDR imagery is applicable to a broad range of scenarios, such as film and television, medical imaging and manufacturing. The information provided by HDR imaging can equal, or even exceed, that which is perceptible by the human visual system (HVS) with, or without, adaption to the ambient illumination [3].

The detail provided by HDR imaging has been the focus of previous studies [9], however traditionally, studies on the use of HDR imagery have been conducted in dark environments [1, 6, 10] to ensure that the provided detail in the imagery is visible. However, there have been no studies presented to date that investigate the amount of detail visible under brighter ambient illuminations. Conditions of up to 700 lx [11] and peak brightnesses of up to 20,000 cd/m^2 [6] have been tested, but not in combination. In environments such as those found outdoors, the ambient illumination can greatly exceed any previously tested.

An experiment in which participants watched HDR video on an HDR display under different ambient illuminations showed that visual fatigue was not a particular problem [11]. There was, however, a significant difference in the choice of brightness and contrast depending on the lighting conditions. The two experiments were conducted on a display with a peak luminance of 4,000 cd/m^2 and a black level of <0.001 cd/m^2 at seven ambient illuminations. <0.01, 0.75, 8.5, 28 and 74 lx in the first and, <0.01, 70 and 700 lx in the second.

Daly et al. [6] conducted an experiment in a dark room that showed for diffuse imagery the average preferred range of display was between 0.1 to 650 cd/m^2. However to satisfy 90% of the population the diffuse range should be between 0.005 to 3,000 cd/m^2 and greater than 20,000 cd/m^2 for highlights [5].

The collected data evaluates both an increased peak brightness and an in brighter ambient illumination than previous studies.

[1] https://www.utad.pt/en/

[2] https://www.inesctec.pt/en/

Application	Typical ambient illumination (lx)
Home theatre [8]	∼0
Advised home environment viewing conditions [8]	200
Inside on a sunny day	3,000
Hotel pool bar	80,000

Table 1: The ambient illuminance of locations for some target applications for HDR displays.

3 Aims/Objectives

The aim of this experiment is to analyse the dynamic range that can be perceived by the HVS under a range of real-world ambient illuminations and display brightnesses. The hypothesis is that in conditions such as outdoors on a sunny day, where the ambient illumination can reach 80,000 lx [4], a higher peak display luminance will allow a greater amount of information in the image to be perceived. To evaluate the dynamic range that can be perceived in the ambient illuminations proposed, a display is required that can output a wide range of luminances. The display chosen for the experiment was a prototype HDR display developed by SIM2 [12]. It was chosen for it's ability of accurately emitting images from $0.05\,\mathrm{cd/m^2}$ to $10{,}000\,\mathrm{cd/m^2}$, no other displays capable of outputting this range of luminance are available.

The motivation behind the experiment and the reason for considering environments covering a large range of lighting conditions was to simulate the wide variety of locations in which HDR displays could be employed in the future. Table 1 shows a selection of typical ambient illuminations at potential locations for HDR displays. Should HDR displays become ubiquitous, it can be expected that they will find application in a wide variety of environments, such as an advertising display in a shopping centre and the bar next to a hotel pool.

4 Data Sharing

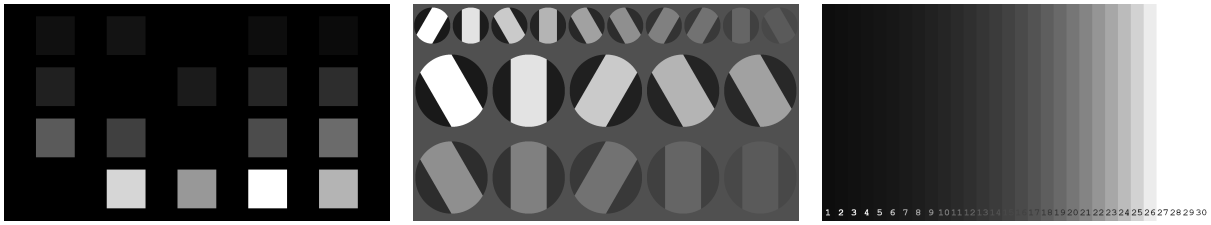
The data was recorded and will be transferred in a pseudonymised format containing the age and gender of the participants along with their responses to the stimuli. The data will be sent in an Excel spreadsheet attached to an email encrypted with SSL and transmitted between the email servers of the UTAD and the University of Warwick. Once received, the pseudonymised data will be stored on University of Warwick computers. Only the primary researcher and supervisors will have access to the data transferred from Portugal.

The primary researcher and their supervisors will analyse the data on University of Warwick computers in the IDL, WMG, central campus. The analysis performed on the data will be repeated measured factorial ANOVA to identify if the dependent variables are affected by the independent variables. Further processing of the data will be use non-linear regression to fit a curve to the data and perform k-fold validation to ensure the fit is robust. The data will be kept in digital format only for 10 years from the date of transfer or until all research is completed using the data, whichever is sooner.

The study will be published in collaboration with the Portuguese University. Researchers from both Warwick and UTAD will be authors on the publication. The anonymised post-analysis data will be available to the researchers from the Portuguese University to allow them to edit the publication. There is no formal link between the researchers at Warwick and Portugal. This is a collaborative effort only.

5 Design/Methodology

To evaluate the amount of detail that can be perceived on a display of a given brightness and in a certain ambient environment, the participants were asked to report the visibility of details on a number of test charts. Each chart had the perceptible details displayed at a range of brightnesses, it is expected that the number of details visible will decrease at the ambient illumination increases. This is due to the adaption of the eye to the environment and therefore no longer being able to perceive details which are more similar



(a) *Squares*: patches separated by 1.5 stops.

(b) *Ayres*: circles with a peak luminance of 20 stops above the background.

(c) *Bars*: luminance ramp containing 30 gradations separated by 1 stop each.

Figure 1: Test charts for use in the proposed experiment. The charts were graded at a range of different luminances, 1,000, 2,000, 4,000 and 8,000 cd/m^2. Here they have been tone-mapped for presentation. *Squares* and *Ayres* show example patterns as the layout was randomised.

to the background shade of the chart. The details will be marked down on paper and the participants will be sitting the experiment in groups.

The experiment comprised of three tasks designed to determine the range visible on displays at simulated luminances under various ambient illuminations. These tasks were as follows:

Squares: participants are asked to count the number of squares;

Ayres: participants are asked to give the orientation of the circles on Ayres charts [2].

Bars: participants are asked to identify the minimum and maximum bars they can discern;

Each task has an associated with a test chart, these are shown in Section 5.1.

These tasks have been selected as they can be conducted efficiently and provide robust results. In addition, the Ayers chart has been used previously to successfully evaluate the dynamic range of an HDR display [9]. The results of the three tasks under different ambient conditions are the dependent variables of the experiment.

The tasks were performed under two varying conditions corresponding to the two independent variables in the experiment. The first independent variable, *brightness*, will be within-participants and is the luminance of the display. Five conditions were considered: 500, 1,000, 2,000, 4,000 and 8,000 cd/m^2. All luminance levels will be simulated on the same display by adjusting the maximum output brightness.

The second independent variable, *ambient*, a between-participants independent variable, represents the ambient illumination in the environment. The conditions: 0, 200, 3,000 and 80,000 lx will be considered. These represent no light, an indoor artificially lit working environment, an indoor sunlit (but not directly) environment, and a sunny outdoor environment respectively. The ambient illumination will be measured using a calibrated luminance meter placed in front of the display and illuminances within 1% of the target outdoor brightness will be considered acceptable. The experiment had a duration of 5 minutes and the luminance measured again at the end to ensure consistency. If the luminance differs from that recorded at the beginning of the condition, the results will be discarded and the condition rerun. Participants will view all *brightness* conditions for a single given *ambient* condition. This consists of 3 tasks at 5 *brightnesses*, each shown for 15 seconds, with a 5 second 100 cd/m^2 grey screen displayed between resulting in the 5 minute duration. The order of the tasks and *brightnesses* was permuted between each set of participants.

5.1 Materials

For each of the three dependent variables, a set of test charts were prepared and graded at a range of luminances, shown in Figure 1. The luminances chosen were 8,000, 4,000, 2,000, 1,000 and 500 cd/m^2, reducing in steps of one stop. The first chart, shown in Figure 1a, contained 16 square luminance patches, the brightest patch shown at the peak luminance of the display with the intensity of each successive patch reduced by 1.25 stops to cover a 20 stop range. The patches were displayed on four rows with four patches on each row along with a single black patch; the order of the patches on each row was randomised so as not to allow the prediction of the location of a patch. The goal of the task was to report the number of visible patches.

The second chart, shown in Figure 1b, contained 20 Ayres circles with varying contrast on a grey background [2], the first row contained 10 smaller circles while the second and third rows contained 5

Parameter	HDR Display
Brand	SIM2 Multimedia
Model	HDR47ES6MB prototype
Form Factor	47”
Static Contrast Ratio	185,000 : 1
Resolution	1,920 × 1,080
Viewing Angle	±85°
Peak Luminance	10,000 cd/m^2

Table 2: Details of the display used in the evaluation.

larger circles. Each was comprised of a dark circle with a light stripe through the middle. Each circle was presented at one of three possible orientations, −30°, 0°, 30°, chosen randomly. The circles were displayed over a range of contrasts, the strip in the highest contrast circle was chosen to be the peak luminance of the chart and was 20 stops lighter than the grey background, the dark region was chosen to be 20 stops darker than the background. Each successive circle has a lightness and darkness 2 stops less than the previous giving 20 stops above and beneath the grey background. The goal of the task was to correctly report the orientation of the circle.

The third chart, shown in Figure 1c, contained 30 full-height bars descending in 1 stop steps from 8,000 cd/m^2. The chart was clamped at the peak luminance required for the grading. Each bar was annotated with a number and the goal of the task was to write down the number corresponding to the first and last perceptually differentiable bar allowing the calculation of the number of stops visible.

The SIM2 display was used at five different luminance levels: 500, 1,000, 2,000, 4,000 and 8,000 cd/m^2 and is capable of a black level of <0.05 cd/m^2. Full details of the display are shown in Table 2.

5.2 Environment

As the purpose of this experiment is to assess how perception is effected by the environment, four proposed environment will be presented.

The environments share common features however, the 47” HDR display will be placed approximately 2 meters from the participants. The participants will be sitting in small groups of five, in two rows so each has an unencumbered view of the display and is within the viewing angle of the display to ensure no loss of luminance information.

The ∼ 0 lx condition will be conducted in a controlled indoor environment with matt black walls, no windows or room lighting, the only illumination will be provided by the display. The 200 lx condition will be conducted in the controlled indoor environment with the adjustable room lighting set to provide the correct ambient illuminance. The 3,000 lx condition will be conducted in a semi-controlled outdoor environment, under shade on a sunny day. Illuminance measurement will be taken with a light meter to ensure the validity of the results. The 80,000 lx condition will be conducted in a semi-controlled outdoor environment, in direct sunlight. A cloudless sunny day will be used to ensure consistent lighting during the duration of the experiment. Illuminance measurement will be taken with a light meter to ensure the validity of the results.

5.3 Procedure

Participants were briefed on the method of the experiment and any questions they might have were answered. Then they were introduced to the environment in which their conditions took place and asked to take a seat, ensuring that they can see the entirety of the display. They were also asked to ensure their vision was corrected to normal if required. Each participant was handed a sheet of paper on a clipboard and a pen in order to report their results. The response questionnaire is shown in Section 9. A brief training run was performed to ensure they were happy with the reporting procedure and understand what was required for each of the test charts. The luminance of the display and illuminance of the environment were noted at this point. Software then displayed the permutation of the images for the experiment on the display and the participants asked to mark down what they can perceive. When all the images have

been displayed the luminance and illuminance were measured again to ensure they remain consistent. If each participant is happy with what they have reported they were thanked for their time and informed that the experiment is complete.

5.4 Participants

To cover all of the conditions, 20 participants were required, however higher numbers would increase the strength of the results upon analysis. Therefore each condition was run twice, and 40 people participated. The minimum age criteria for participant eligibility was 18 years and above and all participants will have normal or corrected to normal vision. In an ideal scenario, the distribution between males and females would be equal and from a wide ranging demographic distribution.

5.4.1 Recruitment strategy

The participants were selected from the Department of Computer Science at the University of Trás-os-Montes and Alto Douro in Vila Real, Portugal. The participants were recruited through announcements after lectures. If the participant agreed to take part in the experiment they were asked to sign a consent form, personal information is only be kept to remove the participant should they withdraw their consent at any stage up to publication.

5.5 Analysis

The aim of this experiment is to determine if the independent variables *brightness* and *ambient* affect the dependent variables, *bars*, *squares* and *Ayres*. The results, therefore, will be analysed with repeated measured factorial ANOVA and corrections such as Greenhouse-Geisser and Bonferroni applied if necessary. It is hoped that the resulting F-value will reject the null hypothesis for a given p-value and therefore the inference can be drawn that the independent variables affect the dependent variables.

6 Ethical Considerations

The experiment will be conducted keeping all ethical considerations in mind including the right to withdraw at any moment during the evaluation. The results of the participant who withdraws from the test will not be retained or used in any way.

6.1 Informed Consent

Informed consent was acquired from each participant before the start of the experiment by means of a thorough briefing including a short training session and a consent letter in addition to a verbal description of the experiments aims and proceedures as in accordance with Portuguese ethical standards.

6.2 Participant confidentiality

The experiment has been designed in such a way that the data analysis will be anonymous and any bias eliminated. Identifying information will only be kept to exercise the right of withdrawal. The evaluation does not require any case studies or personal interviews.

6.3 Data security

Since the experiment does not include any case studies or personal interviews, the data collected will not be sensitive in any way. Additionally, care will be taken such that the collected data does not serve any other interests except for further research in this field. At Warwick, the data will only be accessible to authorised personnel, specifically the student conducting the experiment and the research supervisor.

7 Other Considerations

7.1 Right of withdrawal

No data is stored if the participant decides to withdraw from the experiment midway. In case the participant decides to opt out of the experiment after the completion of the test, their responses will be shredded without further analysis.

7.2 Process of sensitive disclosures

Not applicable.

7.3 Benefits and risks

The psychophysical experiment will be of great benefit to the field of HDR imaging. The perceptibility of HDR imagery is frequently assumed to be uneffected by the ambient environment and little consideration is gone into this aspect of display. The results could well guide the development of future HDR displays and could provide the information required to reduce power consumption while maintaining perceptual quality. The results may also be used to improve the quality of future HDR video compression methods as the amount of information that can be perceived can be predicted at the time of compression.

The risks of participating in this experiment are expected to be minimal. The participant will be asked to sit in a room and view information presented on a display. While the display is brighter than those usually viewed, at 10,000 $\mathrm{cd/m^2}$ the peak brightness of the display is dimmer than a fluorescent lamp [13] and far less than the 8,000,000 $\mathrm{cd/m^2}$ where retinal damage may become possible [7]. If the participant feels uncomfortable viewing the display then they will be required to discontinue the experiment without delay.

References

[1] Ahmet Oğuz Akyüz, Roland Fleming, Bernhard E. Riecke, Erik Reinhard, and Heinrich H. Bülthoff. Do hdr displays support ldr content?: A psychophysical evaluation. *ACM Trans. Graph.*, 26(3), July 2007. ISSN 0730-0301. doi: 10.1145/1276377.1276425.

[2] TJ Ayres. Psychophysical validation of photographic representations. In *The 1996 ASME International Mechanical Engineering Congress and Exposition*, pages 29–33, 1996.

[3] Francesco Banterle, Alessandro Artusi, Kurt Debattista, and Alan Chalmers. *Advanced high dynamic range imaging: theory and practice*. CRC Press, 2011.

[4] Hsi-Hao Chung and Sun Lu. Contrast-ratio analysis of sunlight-readable color lcds for outdoor applications. *Journal of the Society for Information Display*, 11(1):237–242, 3 2003. ISSN 1938-3657. doi: 10.1889/1.1831713.

[5] Scott Daly, Timo Kunkel, Xing Sun, Suzanne Farrell, and Poppy Crum. Viewer preferences for shadow, diffuse, specular, and emissive luminance limits of high dynamic range displays. *SID Symposium Digest of Technical Papers*, 44(1):563–566, 6 2013. ISSN 2168-0159. doi: 10.1002/j.2168-0159.2013.tb06271.x. Distinguished Paper.

[6] Scott J Daly, Timo Kunkel, Xing Sun, Suzanne Farrell, and Poppy Crum. Preference limits of the visual dynamic range for ultra high quality and aesthetic conveyance. In *Human Vision and Electronic Imaging*, page 86510J, 3, 2013.

[7] Donald C Hood and Marcia A Finkelstein. Sensitivity to light. *Handbook of Perception and Human Performance (Vol. 1: Sensory Processes and Perception). John Wiley and Sons, New York.*, 1986.

[8] ITU-R. Bt.2022-0 "general viewing conditions for subjective assessment of quality of sdtv and hdtv television pictures on flat panel displays". *International Telecommunications Union, Geneva*, 8 2012. URL `https://www.itu.int/rec/R-REC-BT.2022-0-201208-I`.

[9] P. Ledda, A. Chalmers, and H. Seetzen. Hdr displays: a validation against reality. In *2004 IEEE International Conference on Systems, Man and Cybernetics (IEEE Cat. No.04CH37583)*, volume 3, pages 2777–2782 vol.3, 10 2004. doi: 10.1109/ICSMC.2004.1400753.

[10] Ratnajit Mukherjee, Kurt Debattista, Thomas Bashford-Rogers, Brian Waterfield, and Alan Chalmers. A study on user preference of high dynamic range over low dynamic range video. *The Visual Computer*, 32(6):825–834, 06 2016. ISSN 1432-2315. doi: 10.1007/s00371-016-1239-7.

[11] Allan G. Rempel, Wolfgang Heidrich, Hiroe Li, and RafałMantiuk. Video viewing preferences for hdr displays under varying ambient illumination. In *Proceedings of the 6th Symposium on Applied Perception in Graphics and Visualization*, APGV '09, pages 45–52, New York, NY, USA, 2009. ACM. ISBN 978-1-60558-743-1. doi: 10.1145/1620993.1621004.

[12] SIM2 Multimedia S.p.A. Sim2 high dynamic range display series. `http://hdr.sim2.it`, July 2015.

[13] Jaroslav Štěpánek. Luminance analysis of linear fluorescent lamps. 2015.

8 Consent Form

Folha de Consentimento

Propósito

O propósito não pode ser divulgado, pois esta informação pode alterar os dados recolhidos. No final da experiência, o sujeito é esclarecido.

Participantes

Os participantes são voluntários que se ofereceram para efetuar a experiência.

Confidencialidade

A informação recolhida apenas é identificada por um "ID" atribuído aleatoriamente.

Procedimento

O procedimento pode variar de teste para teste. As instruções vão sendo dadas ao longo do desenvolvimento das tarefas.

Benefícios

Os participantes não vão receber quaisquer benefícios pela realização da experiência.

Riscos

O procedimento não causa qualquer tipo de lesões e já foi utilizado em estudos previamente realizados.

Compensação

Não está prevista nenhuma compensação. Eventualmente, após a realização da experiência os participantes podem ter acesso a um momento de *debriefing* com acesso gratuito a bebidas (não alcoólicas) e algum tipo de *snack*.

Desistência

A participação neste estudo é completamente voluntária. Os voluntários não estão vinculados a nenhuma obrigação, estes podem terminar o teste quando o desejarem.

Questões ou Dúvidas posteriores

Em caso de dúvidas sobre qualquer um dos tópicos anteriormente descritos, pode enviar um *email* ao investigador para: *mcmelo@inesctec.pt*

Declaração

Fui informado sobre os objetivos deste projeto, assim como dos procedimentos envolvidos nesta experiência. Eu reservo o direito de abandonar a experiência em qualquer fase da mesma e a não ser com o meu consentimento para a sua utilização, toda a informação previamente obtida é destruída ou a minha identidade retirada.

Assinatura: _____

Data: ___/___/___

9 Response Questionnaire

Questionário: Idade:_____ Sexo:_____ Grupo: _____

Imagem 1 – Quantos quadrados consegues contar? _____

Imagem 2 – Identifica qual a orientação da linha dentro do circulo desenhando a linha na tabela abaixo. Na 1ª linha os círculos pequenos na 2ª e 3ª os círculos maiores.

1	2	3	4	5	6	7	8	9	10
1	2	3	4	5					
1	2	3	4	5					

Imagem 3 – Quantos quadrados consegues contar? _____

Imagem 4 – Identifica qual a orientação da linha dentro do circulo desenhando a linha na tabela abaixo. Na 1ª linha os círculos pequenos na 2ª e 3ª os círculos maiores.

1	2	3	4	5	6	7	8	9	10
1	2	3	4	5					
1	2	3	4	5					

Imagem 5 – Quantos quadrados consegues contar? _____

Imagem 6 – Identifica qual a orientação da linha dentro do circulo desenhando a linha na tabela abaixo. Na 1ª linha os círculos pequenos na 2ª e 3ª os círculos maiores.

1	2	3	4	5	6	7	8	9	10
1	2	3	4	5					
1	2	3	4	5					

Imagem 7 – Quantos quadrados consegues contar? _____

Imagem 8 – Identifica qual a orientação da linha dentro do circulo desenhando a linha na tabela abaixo. Na 1ª linha os círculos pequenos na 2ª e 3ª os círculos maiores.

1	2	3	4	5	6	7	8	9	10
1	2	3	4	5					
1	2	3	4	5					

Imagem 9 – Quantos quadrados consegues contar? _____

Imagem 10 – Identifica qual a orientação da linha dentro do circulo desenhando a linha na tabela abaixo. Na 1ª linha os círculos pequenos na 2ª e 3ª os círculos maiores.

1	2	3	4	5	6	7	8	9	10
1	2	3	4	5					
1	2	3	4	5					

PRIVATE

Mr Jonathan Hatchett
WMG
University of Warwick
Coventry
CV4 7AL

09 May 2018

Dear Mr Hatchett,

Study Title and BSREC Reference: *Psychophysical Evaluation of Perceptible Detail in Natural Environments* REGO-2018-2184

Thank you for submitting the above-named project to the University of Warwick Biomedical and Scientific Research Ethics Committee for research ethical review.

I am pleased to advise that research ethical approval is granted.

In undertaking your study, you are required to comply with the University of Warwick's *Research Data Management Policy*, details of which may be found on the Research and Impact Services' webpages, under "Codes of Practice & Policies" » "Research Code of Practice" » "Data & Records" » "Research Data Management Policy", at:
http://www2.warwick.ac.uk/services/ris/research_integrity/code_of_practice_and_policies/research_code_of_practice/datacollection_retention/research_data_mgt_policy

You are also required to comply with the University of Warwick's *Information Classification and Handling Procedure*, details of which may be found on the University's Governance webpages, under "Governance" » "Information Security" » "Information Classification and Handling Procedure", at:
http://www2.warwick.ac.uk/services/gov/informationsecurity/handling.
Investigators should familiarise themselves with the classifications of information defined therein, and the requirements for the storage and transportation of information within the different classifications:

> *Information Classifications:*
> http://www2.warwick.ac.uk/services/gov/informationsecurity/handling/classifications
> *Handling Electronic Information:*
> http://www2.warwick.ac.uk/services/gov/informationsecurity/handling/electronic/
> *Handling Paper or other media*
> http://www2.warwick.ac.uk/services/gov/informationsecurity/handling/paper/.

Please also be aware that BSREC grants **ethical approval** for studies. **The seeking and obtaining of *all* other necessary approvals is the responsibility of the investigator**

These other approvals may include, but are not limited to:

1. Any necessary agreements, approvals, or permissions required in order to comply with the University of Warwick's Financial Regulations and Procedures.
2. Any necessary approval or permission required in order to comply with the University of Warwick's Quality Management System and Standard Operating Procedures for the governance, acquisition, storage, use, and disposal of human samples for research.
3. All relevant University, Faculty, and Divisional/Departmental approvals, if an employee or student of the University of Warwick.
4. Approval from the applicant's academic supervisor and course/module leader (as appropriate), if a student of the University of Warwick.
5. Health Research Authority (HRA) approval for research studies undertaken in NHS Trusts (previously NHS R+D approval).
6. NHS Trust Clinical Audit Approval, for clinical audit studies undertaken in NHS Trusts.
7. Approval from Departmental or Divisional Heads, as required under local procedures, within Health and Social Care organisations hosting the study.
8. Local ethical approval for studies undertaken overseas, or in other HE institutions in the UK.
9. Approval from Heads (or delegates thereof) of UK Medical Schools, for studies involving medical students as participants.
10. Permission from Warwick Medical School to access medical students or medical student data for research or evaluation purposes.
11. NHS Trust Caldicott Guardian Approval, for studies where identifiable data is being transferred outside of the direct clinical care team. Individual NHS Trust procedures vary in their implementation of Caldicott guidance, and local guidance must be sought.
12. Any other approval required by the institution hosting the study, or by the applicant's employer.

There is no requirement to supply documentary evidence of any of the above to BSREC, but applicants should hold such evidence in their Study Master File for University of Warwick auditing and monitoring purposes. You may be required to supply evidence of any necessary approvals to other University functions, e.g. The Finance Office, Research & Impact Services (RIS), or your Department/School.

May I take this opportunity to wish you success with your study, and to remind you that any Substantial Amendments to your study require approval from BSREC before they may be implemented.

Yours sincerely

pp. [redacted]

Dr David Ellard
Chair
Biomedical and Scientific
Research Ethics Sub-Committee

Biomedical and Scientific
Research Ethics Sub-Committee
Research & Impact Services
University of Warwick
Coventry, CV4 8UW.
E: BSREC@Warwick.ac.uk

http://www2.warwick.ac.uk/services/ris/research_integrity/researchethicscommittees/biomed

References

[AFR+07] Ahmet Oğuz Akyüz, Roland Fleming, Bernhard E. Riecke, Erik Reinhard and Heinrich H. Bülthoff. Do hdr displays support ldr content?: a psychophysical evaluation. *ACM Trans. Graph.*, 26(3), July 2007. ISSN: 0730-0301. DOI: 10.1145/1276377.1276425.

[App68] Arthur Appel. Some techniques for shading machine renderings of solids. In *Proceedings of the April 30–May 2, 1968, Spring Joint Computer Conference*, AFIPS '68 (Spring), pages 37–45, Atlantic City, New Jersey. ACM, 1968. DOI: 10.1145/1468075.1468082. URL: http://doi.acm.org/10.1145/1468075.1468082.

[Arr17] Arri Group. Alexa tech specs, September 2017. URL: http://www.arri.com/camera/alexa/cameras/camera_details/alexa-classic-ev/subsection/technical_data/.

[AMR+15] Alessandro Artusi, Rafał K. Mantiuk, Thomas Richter, Philippe Hanhart, Pavel Korshunov, Massimiliano Agostinelli, Arkady Ten and Touradj Ebrahimi. Overview and evaluation of the jpeg xt hdr image compression standard. *Journal of Real-Time Image Processing*, December 2015. ISSN: 1861-8219. DOI: 10.1007/s11554-015-0547-x.

[AR+98] National Electrical Manufacturers Association, American College of Radiology et al. *Digital imaging and communications in medicine (DICOM), Part 14: Grayscale Standard Display Function*. National Electrical Manufacturers Association, 1998.

[AMS08] Tunç O Aydın, Rafal Mantiuk and Hans-Peter Seidel. Extending quality metrics to full luminance range images. In *Electronic Imaging 2008*, 68060B–68060B. International Society for Optics and Photonics, 2008.

[Ayr96] TJ Ayres. Psychophysical validation of photographic representations. In *The 1996 ASME International Mechanical Engineering Congress and Exposition*, pages 29–33, 1996.

[BAD+11] Francesco Banterle, Alessandro Artusi, Kurt Debattista and Alan Chalmers. *Advanced high dynamic range imaging: theory and practice*. CRC Press, 2011.

[BAD+10] Francesco Banterle, Alessandro Artusi, Kurt Debattista, Patrick Ledda, Alan Chalmers, Gavin John Edwards and Gerhard Bonnet. HDR video data compression devices and methods, January 2010. EP Patent 2,144,444.

[BDA+09] Francesco Banterle, Kurt Debattista, Alessandro Artusi, Sumanta Pattanaik, Karol Myszkowski, Patrick Ledda and Alan Chalmers. High dynamic range imaging and low dynamic range expansion for generating hdr content. In *Computer graphics forum*, volume 28 of number 8, pages 2343–2367. Wiley Online Library, October 2009.

[Bar92] Peter GJ Barten. Physical model for the contrast sensitivity of the human eye. In *SPIE/IS&T 1992 Symposium on Electronic Imaging: Science and Technology*, pages 57–72. International Society for Optics and Photonics, 1992.

[BET+08] Herbert Bay, Andreas Ess, Tinne Tuytelaars and Luc Van Gool. Speeded-up robust features (surf). *Computer vision and image understanding*, 110(3):346–359, 2008.

[BCM+17] Cambodge Bist, Rémi Cozot, Gérard Madec and Xavier Ducloux. Tone expansion using lighting style aesthetics. *Computers & Graphics*, 62:77–86, 2017. ISSN: 0097-8493. DOI: https://doi.org/10.1016/j.cag.2016.12.006.

[Bjø01] Gisle Bjøntegaard. Calculation of average psnr differences between rd-curves. *Doc. VCEG-M33 ITU-T Q6/16, Austin, TX, USA, 2-4 April 2001*, 2001.

[BC15] T. Borer and A. Cotton. A "display independent" high dynamic range television system. *IBC2015*, 1st September 2015.

[Bor14] Tim Borer. Non-linear opto-electrical transfer functions for high dynamic range television, 2014.

[Bro70] Charles George Broyden. The convergence of a class of double-rank minimization algorithms 1. general considerations. *IMA Journal of Applied Mathematics*, 6(1):76–90, 1970.

[Can14] Canon Inc. Canon EOS 5D Mark III - specifications, September 2014. URL: http://www.canon.co.uk/For_Home/Product_Finder/Cameras/Digital_SLR/EOS_5D_Mark_III/#p-specification.

[CMH+15] Alan Chalmers, Joshua McNamee, Jonathan Hatchett, Ratnajit Mukherjee, Igor Olaizola and Kurt Debattista. 12 bits is simply not enough for HDR video! *NAB*, 2015.

[CKL14] Hojin Cho, Seon Joo Kim and Seungyong Lee. Single-shot high dynamic range imaging using coded electronic shutter. *Computer Graphics Forum*, 33(7):329–338, 2014. ISSN: 1467-8659. DOI: 10.1111/cgf.12501.

[CL03] Hsi-Hao Chung and Sun Lu. Contrast-ratio analysis of sunlight-readable color lcds for outdoor applications. *Journal of the Society for Information Display*, 11(1):237–242, March 2003. ISSN: 1938-3657. DOI: 10.1889/1.1831713.

[DKS+13a] Scott J Daly, Timo Kunkel, Xing Sun, Suzanne Farrell and Poppy Crum. Preference limits of the visual dynamic range for ultra high quality and aesthetic conveyance. In *Human Vision and Electronic Imaging*, 86510J, 3, 2013.

[DKS+13b] Scott Daly, Timo Kunkel, Xing Sun, Suzanne Farrell and Poppy Crum. Viewer preferences for shadow, diffuse, specular, and emissive luminance limits of high dynamic range displays. *SID Symposium Digest of Technical Papers*, 44(1):563–566, June 2013. ISSN: 2168-0159. DOI: 10.1002/j.2168-0159.2013.tb06271.x. Distinguished Paper.

[DBS+15] Kurt Debattista, Thomas Bashford-Rogers, Elmedin Selmanović, Ratnajit Mukherjee and Alan Chalmers. Optimal exposure compression for high dynamic range content. *The Visual Computer*, 31(6):1089–1099, May 2015. ISSN: 1432-2315. DOI: 10.1007/s00371-015-1121-z.

[DM97] Paul E. Debevec and Jitendra Malik. Recovering high dynamic range radiance maps from photographs. In *Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '97, pages 369–378, New York, NY, USA. ACM Press/Addison-Wesley Publishing Co., 1997. ISBN: 0-89791-896-7. DOI: 10.1145/258734.258884.

[DSC14] DSC Labs. Xyla® ultra high dynamic range charts, September 2014. URL: http://dsclabs.com/dsc-products/rear-lit/xyla/.

[EF98] Fritz Ebner and Mark D. Fairchild. Development and testing of a color space (ipt) with improved hue uniformity. *Color and Imaging Conference*:8–13, 1998. ISSN: 2166-9635.

[EMU16] G. Eilertsen, R. K. Mantiuk and J. Unger. A high dynamic range video codec optimized by large-scale testing. In *2016 IEEE International Conference on Image Processing (ICIP)*, pages 1379–1383, September 2016. DOI: 10.1109/ICIP.2016.7532584.

[ED04] Elmar Eisemann and Frédo Durand. Flash photography enhancement via intrinsic relighting. In *ACM transactions on graphics (TOG)*, volume 23 of number 3, pages 673–678. ACM, 2004.

[Fai07] Mark D Fairchild. The hdr photographic survey. In *Color and Imaging Conference*, volume 2007 of number 1, pages 233–238. Society for Imaging Science and Technology, 2007.

[Fec38] G. T. Fechner. Ueber eine scheibe zur erzeugung subjectiver farben. *Annalen der Physik*, 121(10):227–232, 1838. ISSN: 1521-3889. DOI: 10.1002/andp.18381211004.

[FPS+96] James A Ferwerda, Sumanta N Pattanaik, Peter Shirley and Donald P Greenberg. A model of visual adaptation for realistic image synthesis. In *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*, pages 249–258. ACM, 1996.

[FB81] Martin A. Fischler and Robert C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM*, 24(6):381–395, June 1981. ISSN: 0001-0782. DOI: 10.1145/358669.358692. URL: http://doi.acm.org/10.1145/358669.358692.

[Fle70] Roger Fletcher. A new approach to variable metric algorithms. *The computer journal*, 13(3):317–322, January 1970.

[FGE+14] Jan Froehlich, Stefan Grandinetti, Bernd Eberhardt, Simon Walter, Andreas Schilling and Harald Brendel. Creating cinematic wide gamut hdr-video for the evaluation of tone mapping operators and hdr-displays. In 2014.

[GGC+09] O. Gallo, N. Gelfand, W. Chen, M. Tico and K. Pulli. Artifact-free high dynamic range imaging. *IEEE International Conference on Computational Photography (ICCP)*, April 2009.

[GT11] Jens-Uwe Garbas and Herbert Thoma. Temporally coherent luminance-to-luma mapping for high dynamic range video coding with h.264/avc. In *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*, pages 829–832. IEEE, 2011.

[GBB11] Xavier Glorot, Antoine Bordes and Yoshua Bengio. Deep sparse rectifier neural networks. In Geoffrey Gordon, David Dunson and Miroslav Dudík, editors, *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, volume 15 of *Proceedings of Machine Learning Research*, pages 315–323, Fort Lauderdale, FL, USA. PMLR, November 2011. URL: http://proceedings.mlr.press/v15/glorot11a.html.

[Gol70] Donald Goldfarb. A family of variable-metric methods derived by variational means. *Mathematics of computation*, 24(109):23–26, 1970.

[GRG+17] D. Gommelet, A. Roumy, C. Guillemot, M. Ropert and J. L. Tanou. Gradient-based tone mapping for rate-distortion optimized backward-compatible high dynamic range compression. *IEEE Transactions on Image Processing*, PP(99):1–1, 15th August 2017. ISSN: 1057-7149. DOI: 10.1109/TIP.2017.2740159.

[GKT+13] Miguel Granados, Kwang In Kim, James Tompkin and Christian Theobalt. Automatic noise modeling for ghost-free hdr reconstruction. *ACM Trans. Graph.*, 32(6):201:1–201:10, November 2013. ISSN: 0730-0301. DOI: 10.1145/2508363.2508410. URL: http://doi.acm.org/10.1145/2508363.2508410.

[GRH16] Adrian Grange, Peter de Rivaz and Jonathan Hunt. Vp9 bitstream & decoding process specification. *Google*, March 2016.

[Gro06] Thorsten Grosch. Fast and robust high dynamic range image generation with camera and object movement. *Vision, Modeling and Visualization, RWTH Aachen*:277–284, 2006.

[GN02] Michael D. Grossberg and Shree K. Nayar. What can be known about the radiometric response from images? In *Proceedings of the 7th European Conference on Computer Vision-Part IV*, ECCV '02, pages 189–205, London, UK. Springer-Verlag, 2002. ISBN: 3-540-43748-7. URL: http://dl.acm.org/citation.cfm?id=645318.649359.

[HSM+00] Richard HR Hahnloser, Rahul Sarpeshkar, Misha A Mahowald, Rodney J Douglas and H Sebastian Seung. Digital selection and analogue amplification coexist in a cortex-inspired silicon circuit. *Nature*, 405(6789):947, June 2000.

[HBP+15] Philippe Hanhart, Marco V Bernardo, Manuela Pereira, António MG Pinheiro and Touradj Ebrahimi. Benchmarking of objective quality metrics for hdr image quality assessment. *EURASIP Journal on Image and Video Processing*, 2015(1):1–18, 2015.

[HDR17] HDRC. Camcube, September 2017. URL: http://www.ims-chips.de/content/pdftext/HDRC_Imager_Camera_Feature3.pdf.

[HLL+11] Yong Seok Heo, Kyoung Mu Lee, Sang Uk Lee, Youngsu Moon and Joonhyuk Cha. Ghost-free high dynamic range imaging. In *Proceedings of the 10th Asian Conference on Computer Vision - Volume Part IV*, ACCV'10, pages 486–500, Queenstown, New Zealand. Springer-Verlag, 2011. ISBN: 978-3-642-19281-4. URL: http://dl.acm.org/citation.cfm?id=1966111.1966151.

[Hub12] Ian Hubert. Tears of steel, 2012.

[IEE85] IEEE. Standard for binary floating-point arithmetic. *ANSI/IEEE Std 754-1985*, March 1985. DOI: 10.1109/IEEESTD.1985.82928.

[IEE08] IEEE. Standard for floating-point arithmetic. *IEEE Std 754-2008*:1–70, August 2008. DOI: 10.1109/IEEESTD.2008.4610935.

[Ind14] Industrial Light & Magic. OpenEXR, October 2014. URL: http://www.openexr.com/.

[ITU11] ITU-R. Bt.601-7: "studio encoding parameters of digital television for standard 4:3 and wide screen 16:9 aspect ratios". *International Telecommunications Union, Geneva*, March 2011. URL: https://www.itu.int/rec/R-REC-BT.601-7-201103-I/en.

[ITU12] ITU-R. Bt.2022-0 "general viewing conditions for subjective assessment of quality of sdtv and hdtv television pictures on flat panel displays". *International Telecommunications Union, Geneva*, August 2012. URL: https://www.itu.int/rec/R-REC-BT.2022-0-201208-I.

[ITU15] ITU-R. Bt.2020-2 "parameter values for ultra-high definition television systems for production and international programme exchange". *International Telecommunications Union, Geneva*, October 2015. URL: https://www.itu.int/rec/R-REC-BT.2020-2-201510-I.

[ITU16] ITU-R. Bt.2100-1 "image parameter values for high dynamic range television for use in production and international programme exchange". *International Telecommunications Union, Geneva*, June 2016. URL: https://www.itu.int/rec/R-REC-BT.2100-1-201706-I/en.

[JWL05] Katrien Jacobs, Greg Ward and Celine Loscos. Automatic hdri generation of dynamic environments. In *ACM SIGGRAPH 2005 Sketches*, SIGGRAPH '05, Los Angeles, California. ACM, 2005. DOI: 10.1145/1187112.1187163. URL: http://doi.acm.org/10.1145/1187112.1187163.

[JPR14] Garima Jain, Anand Plappally and Shanmuganathan Raman. Internethdr: enhancing an ldr image using visually similar internet images. In *Communications (NCC), 2014 Twentieth National Conference on*, pages 1–6. IEEE, 2014.

[JNT+13] W. Jia, A. Ninan, A. Ten and G.J. Ward. Encoding, decoding, and representing high dynamic range images, November 2013. URL: `https://www.google.com/patents/US20130294689`. US Patent App. 13/938,795.

[KUW+03] Sing Bing Kang, Matthew Uyttendaele, Simon Winder and Richard Szeliski. High dynamic range video. In *ACM Transactions on Graphics (TOG)*, volume 22 of number 3, pages 319–325. ACM, 2003.

[Kel77] D. H. Kelly. Visual contrast sensitivity. *Optica Acta: International Journal of Optics*, 24(2):107–129, 1977. DOI: `10.1080/713819495`.

[KB14] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *CoRR*, abs/1412.6980, December 2014. URL: `http://arxiv.org/abs/1412.6980`.

[KGB+14] Joel Kronander, Stefan Gustavson, Gerhard Bonnet, Anders Ynnerman and Jonas Unger. A unified framework for multi-sensor hdr video reconstruction. *Signal Processing: Image Communication*, 29(2):203–215, 2014.

[LRW+98] Jeffrey C Lagarias, James A Reeds, Margaret H Wright and Paul E Wright. Convergence properties of the nelder–mead simplex method in low dimensions. *SIAM Journal on optimization*, 9(1):112–147, 1998.

[Lar98] Gregory Ward Larson. Logluv encoding for full-gamut, high-dynamic range images. *Journal of Graphics Tools*, 3(1):15–31, March 1998. ISSN: 1086-7651. DOI: `10.1080/10867651.1998.10487485`.

[LLF13] Sébastien Lasserre, Fabrice LeLéannec and Edouard Francois. Description of hdr sequences proposed by technicolor. *ISO/IEC JTC1/SC29/WG11 JCTVC-P0228], IEEE, San Jose, USA*, 2013.

[LBH15] Yann LeCun, Yoshua Bengio and Geoffrey Hinton. Deep learning. *Nature*, 521(7553):436–444, February 2015.

[LCS04] P. Ledda, A. Chalmers and H. Seetzen. Hdr displays: a validation against reality. In *2004 IEEE International Conference on Systems, Man and Cybernetics (IEEE Cat. No.04CH37583)*, volume 3, 2777–2782 vol.3, October 2004. DOI: `10.1109/ICSMC.2004.1400753`.

[LK08] Chul Lee and Chang-Su Kim. Rate-distortion optimized compression of high dynamic range videos. In *Proceedings of the 16th European Signal Processing Conference*, 2008.

[LPY+16a] Taoran Lu, Fangjun Pu, Peng Yin, Tao Chen, Walt Husak, Jaclyn Pytlarz, Robin Atkins, Jan Fr-hlich and Guan-Ming Su. Itp colour space and its compression performance for high dynamic range and wide colour gamut video distribution. *ZTE Communications*, 14, 1, February 2016.

[LPY+16b] Taoran Lu, Fangjun Pu, Peng Yin, Jaclyn Pytlarz, Tao Chen and Walt Husak. Adaptive reshaper for high dynamic range and wide color gamut video compression. In volume 9971, pages 1–14, September 2016. DOI: 10.1117/12.2236050.

[LK81] Bruce D. Lucas and Takeo Kanade. An iterative image registration technique with an application to stereo vision. In *Proceedings of the 7th International Joint Conference on Artificial Intelligence - Volume 2*, IJCAI'81, pages 674–679, Vancouver, BC, Canada. Morgan Kaufmann Publishers Inc., 1981. URL: http://dl.acm.org/citation.cfm?id=1623264.1623280.

[LFH14] Ajay Luthra, Edouard François and Walt Husak. Requirements and use cases for hdr and wcg content distribution, 2014. ISO/IEC JTC 1/SC 29/WG 11 (MPEG) Doc.

[LFH15] Ajay Luthra, Edouard Franois and Walt Husak. Call for evidence (cfe) for hdr and wcg video coding. *ISO/IEC JTC1/SC29/WG11 MPEG, Geneva, Switzerland, Tech. Rep N*, 15083, February 2015.

[Mag14] Magic Lantern. Magic lantern canon firmware, October 2014. URL: http://www.magiclantern.fm/.

[MMM+11] Z. Mai, H. Mansour, R. Mantiuk, P. Nasiopoulos, R. Ward and W. Heidrich. Optimizing a tone curve for backward-compatible high dynamic range image and video compression. *IEEE Transactions on Image Processing*, 20(6):1558–1571, June 2011. ISSN: 1057-7149. DOI: 10.1109/TIP.2010.2095866.

[MP94] S Mann and RW Picard. "Being undigital" with digital cameras. *MIT Media Lab Perceptual Computing Technical Report*, 1994.

[MEM+06] Rafał Mantiuk, Alexander Efremov, Karol Myszkowski and Hans-Peter Seidel. Backward compatible high dynamic range mpeg video compression. In *ACM Transactions on Graphics (TOG)*, volume 25 of number 3, pages 713–723. ACM, 2006.

[MKR+11] Rafał Mantiuk, Kil Joong Kim, Allan G Rempel and Wolfgang Heidrich. HDR-VDP-2: a calibrated visual metric for visibility and quality predictions in all luminance conditions. In *ACM Transactions on Graphics (TOG)*, volume 30 of number 4, page 40. ACM, 2011.

[MKM+04] Rafał Mantiuk, Grzegorz Krawczyk, Karol Myszkowski and Hans-Peter Seidel. Perception-motivated high dynamic range video encoding. In *ACM Transactions on Graphics (TOG)*, volume 23 of number 3, pages 733–741. ACM, 2004.

[MMS06] Rafał Mantiuk, Karol Myszkowski and Hans-Peter Seidel. Lossy compression of high dynamic range images and video. In *Electronic Imaging 2006*, pages 60570V–60570V. International Society for Optics and Photonics, 2006.

[MBB+15] Miguel Melo, Maximino Bessa, Luís Barbosa, Kurt Debattista and Alan Chalmers. Screen reflections impact on hdr video tone mapping for mobile devices: an evaluation study. *EURASIP Journal on Image and Video Processing*, 2015(1):44, December 2015. ISSN: 1687-5281. DOI: 10.1186/s13640-015-0094-1.

[MND13] Scott Miller, Mahdi Nezamabadi and Scott Daly. Perceptual signal coding for more efficient usage of bit codes. In *SMPTE Conferences*, volume 2013 of number 10, pages 1–9. Society of Motion Picture and Television Engineers, May 2013.

[MN99] Tomoo Mitsunaga and Shree K Nayar. Radiometric self calibration. In *Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on.* Volume 1. IEEE, 1999.

[MT10] Ajit Motra and Herbert Thoma. An adaptive logluv transform for high dynamic range video compression. In *Image Processing (ICIP), 2010 17th IEEE International Conference on*, pages 2061–2064. IEEE, 2010.

[MBG+13] D. Mukherjee, J. Bankoski, A. Grange, J. Han, J. Koleszar, P. Wilkins, Y. Xu and R. Bultje. The latest open-source video codec vp9 - an overview and preliminary results. In *2013 Picture Coding Symposium (PCS)*, pages 390–393, December 2013. DOI: 10.1109/PCS.2013.6737765.

[MDB+16a] R. Mukherjee, K. Debattista, T. Bashford-Rogers, P. Vangorp, R. Mantiuk, M. Bessa, B. Waterfield and A. Chalmers. Objective and subjective evaluation of high dynamic range video compression. *Signal Processing: Image Communication*, 47:426–437, September 2016. ISSN: 0923-5965. URL: http://www.sciencedirect.com/science/article/pii/S0923596516301084.

[MDB+16b] Ratnajit Mukherjee, Kurt Debattista, Thomas Bashford-Rogers, Brian Waterfield and Alan Chalmers. A study on user preference of high dynamic range over low dynamic range video. *The Visual Computer*, 32(6):825–834, June 2016. ISSN: 1432-2315. DOI: 10.1007/s00371-016-1239-7.

[NR66] Ken-Ichi Naka and William AH Rushton. S-potentials from luminosity units in the retina of fish (cyprinidae). *The Journal of physiology*, 185(3):587–599, January 1966.

[NM00] Shree K Nayar and Tomoo Mitsunaga. High dynamic range imaging: spatially varying pixel exposures. In *Computer Vision and Pattern Recognition, 2000. Proceedings. IEEE Conference on*, volume 1, pages 472–479. IEEE, 2000.

[NM65] J. A. Nelder and R. Mead. A simplex method for function minimization. *The Computer Journal*, 7(4):308–313, January 1965. DOI: 10.1093/comjnl/7.4.308.

[Pan17a] Panasonic. Tx-65dx902b, September 2017. URL: http://www.panasonic.com/uk/consumer/televisions/4KTV/tx-65dx902b.html.

[Pan17b] Panasonic. Tx-65ez1002b, September 2017. URL: http://www.panasonic.com/uk/consumer/televisions/4KTV/tx-65ez1002b.html.

[Poy93] Charles A Poynton. Smpte tutorial:"gamma" and its disguises: the nonlinear mappings of intensity in perception, crts, film, and video. *SMPTE journal*, 102(12):1099–1108, 1993.

[RB15] Association of Radio Industries and Businesses. STD-B67: essential parameter values for the extended image dynamic range television (EIDRTV) system for programme production, July 2015.

[Red14] Red Digital Cinema Camera Company. Epic tech specs, October 2014. URL: http://www.red.com/products/epic#tech-specs.

[RHD+10] Erik Reinhard, Wolfgang Heidrich, Paul Debevec, Sumanta Pattanaik, Greg Ward and Karol Myszkowski. *High dynamic range imaging: acquisition, display, and image-based lighting*. Morgan Kaufmann, 2010.

[RHL+09] Allan G. Rempel, Wolfgang Heidrich, Hiroe Li and Rafał Mantiuk. Video viewing preferences for hdr displays under varying ambient illumination. In *Proceedings of the 6th Symposium on Applied Perception in Graphics and Visualization*, APGV '09, pages 45–52, Chania, Crete, Greece. ACM, 2009. ISBN: 978-1-60558-743-1. DOI: 10.1145/1620993.1621004.

[Sam16] Samsung. Js9800, September 2016. URL: http://www.samsung.com/hk_en/tvs/suhd-js9800/UA65JS9800JXZK/.

[Sch96] Paul Scheunders. A genetic lloyd-max image quantization algorithm. *Pattern Recognition Letters*, 17(5):547–556, 1996.

[SHS+04] Helge Seetzen, Wolfgang Heidrich, Wolfgang Stuerzlinger, Greg Ward, Lorne Whitehead, Matthew Trentacoste, Abhijeet Ghosh and Andrejs Vorozcovs. High dynamic range display systems. In *ACM Transactions on Graphics (TOG)*, volume 23 of number 3, pages 760–768. ACM, 2004.

[SWW03] Helge Seetzen, Lorne A Whitehead and Greg Ward. 54.2: a high dynamic range display using low and high resolution modulators. In *SID Symposium Digest of Technical Papers*, volume 34 of number 1, pages 1450–1453. Wiley Online Library, 2003.

[Sel13] Elmedin Selmanovic. *Stereoscopic High Dynamic Range Imaging*. PhD thesis, University of Warwick, 2013.

[SYD87] M. I. Sezan, K. l. Yip and S. J. Daly. Uniform perceptual quantization: applications to digital radiography. *IEEE Transactions on Systems, Man, and Cybernetics*, 17(4):622–634, July 1987. ISSN: 0018-9472. DOI: 10.1109/TSMC.1987.289352.

[Sha70] David F Shanno. Conditioning of quasi-newton methods for function minimization. *Mathematics of computation*, 24(111):647–656, 1970.

[SWD05] Gaurav Sharma, Wencheng Wu and Edul N Dalal. The ciede2000 color-difference formula: implementation notes, supplementary test data, and mathematical observations. *Color Research & Application*, 30(1):21–30, 2005.

[SPS+09] Desire Sídíbe, William Puech, Olivier Strauss et al. Ghost detection and removal in high dynamic range images, 2009.

[SIM15] SIM2 Multimedia S.p.A. Sim2 high dynamic range display series. http://hdr.sim2.it, July 2015.

[SMP14] SMPTE. *ST2084:2014 : High Dynamic Range Electro-Optical Transfer Function of Mastering Reference Displays*. Society of Motion Picture and Television Engineers, August 2014. DOI: 10.5594/s9781614828297.

[Son14] Sony. F65 tech specs, October 2014. URL: http://www.sony.co.uk/pro/product/broadcast-products-camcorders-digital-motion-picture-camera/f65/specifications/#specifications.

[Son17a] Sony. Kd-65a1, September 2017. URL: https://www.sony.co.uk/electronics/televisions/a1-series.

[Son17b] Sony. Kd-65zd9bu, September 2017. URL: https://www.sony.co.uk/electronics/televisions/zd9-series/specifications.

[Sph14] SpheronVR Ag. Spheron SceneCam, October 2014. URL: https://www.spheron.com/fileadmin/media/04_Media/Spheron_Camera-SceneCam.pdf.

[SS12] Abhilash Srikantha and Désiré Sidibé. Ghost detection and removal for high dynamic range images: recent advances. *Signal Processing: Image Communication*, 27(6):650–662, 2012.

[SOH+12] Gary J Sullivan, J-R Ohm, Woo-Jin Han and Thomas Wiegand. Overview of the high efficiency video coding (hevc) standard. *Circuits and Systems for Video Technology, IEEE Transactions on*, 22(12):1649–1668, 2012.

[Tor89] Virginia Joanne Torczon. *Multidirectional search: a direct search algorithm for parallel machines*. PhD thesis, Rice University, May 1989.

[TR93] Jack Tumblin and Holly Rushmeier. Tone reproduction for realistic images. *Computer Graphics and Applications, IEEE*, 13(6):42–48, 1993.

[Wal92] Gregory K Wallace. The jpeg still picture compression standard. *IEEE transactions on consumer electronics*, 38(1):xviii–xxxiv, 1992.

[War11] Greg Ward Larson. *Radiance File Formats*. Lawrence Berkeley National Laboratory. November 2011.

[War91] Greg Ward. Real pixels. *Graphics Gems II*:80–83, 1991.

[War03] Greg Ward. Fast, robust image registration for compositing high dynamic range photographs from hand-held exposures. *Journal of graphics tools*, 8(2):17–30, 2003.

[WS04] Greg Ward and Maryann Simmons. Subband encoding of high dynamic range imagery. In *Proceedings of the 1st Symposium on Applied Perception in Graphics and Visualization*, pages 83–90. ACM, 2004.

[WS06] Greg Ward and Maryann Simmons. Jpeg-hdr: a backwards-compatible, high dynamic range extension to jpeg. In *ACM SIGGRAPH 2006 Courses*, SIGGRAPH '06, Boston, Massachusetts. ACM, 2006. ISBN: 1-59593-364-6. DOI: 10.1145/1185657.1185685.

[War94] Gregory J Ward. The radiance lighting simulation and rendering system. In *Proceedings of the 21st annual conference on Computer graphics and interactive techniques*, pages 459–472. ACM, 1994.

[Web34] Ernst Heinrich Weber. *De Pulsu, resorptione, auditu et tactu: Annotationes anatomicae et physiologicae*. CF Koehler, 1834.

[Wei14] Weiss Ag. Civetta 360, October 2014. URL: http://www.weiss-ag.org/home/civetta/.

[WSB+03] Thomas Wiegand, Gary J Sullivan, Gisle Bjøntegaard and Ajay Luthra. Overview of the h. 264/avc video coding standard. *Circuits and Systems for Video Technology, IEEE Transactions on*, 13(7):560–576, 2003.

[ZRB11] Yang Zhang, Erik Reinhard and David Bull. Perception-based high dynamic range video compression with optimal bit-depth transformation. In *Image Processing (ICIP), 2011 18th IEEE International Conference on*, pages 1321–1324. IEEE, 2011.